Structural Analysis of Heterodimeric and Homooligomeric Protein Complexes

by 4-D Fast NMR

by

Su Wang

Department of Biochemistry
Duke University

Date:_____

Approved:

_____
Pei Zhou, Supervisor

_____
Seok-Yong Lee

_____
Leonard Spicer

_____
Tso-Pang Yao

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2014

ABSTRACT

Structural Analysis of Heterodimeric and Homooligomeric Protein Complexes

by 4-D Fast NMR

by

Su Wang

Department of Biochemistry
Duke University

Date:_____
Approved:

_____
Pei Zhou, Supervisor

_____
Seok-Yong Lee

_____
Leonard Spicer

_____
Tso-Pang Yao

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biochemistry in the Graduate School of
Duke University

2014

Abstract

A molecular depiction of the assembly, interaction and regulation of protein complexes is essential to the understanding of biological functions of protein complexes. Structural analysis of protein complexes by Nuclear Magnetic Resonance (NMR) has relied heavily on the detection and assignment of intermolecular Nuclear Overhauser Effects (NOEs) that define the interactions of protons at the molecular interface. Intermolecular NOEs have traditionally been detected from 3-D half-filtered NOE experiments by suppressing intramolecular NOEs prior to NOE transfer. However, due to insufficient suppression of undesirable signals and a lack of dispersion in the H dimension, data analysis is complicated by the interference of residual intramolecular NOEs and assignment ambiguity, both of which can lead to distorted or even erroneously packed protein complex structures. Leveraging the recent development of fast NMR technology based on sparse sampling in our lab, we developed a strategy for reliable identification and assignment of intermolecular NOEs using high resolution 4-D NOE difference spectroscopy. Spectral subtraction of individually labeled components from a uniformly labeled protein complex yields an "omit" spectrum containing only intermolecular NOEs with little signal degeneracy.

The benefit of such a strategy is first demonstrated in structural analysis of a homooligomeric protein complexes, the foldon trimer. We show that intermolecular NOEs collected from the 4-D omit NOE spectrum can be directly utilized for automated

structural analysis of the foldon trimer by CYANA, whereas intermolecular NOEs derived from 3-D half-filtered NOE experiments failed to generate a converged structure under the same condition.

Such a strategy was further demonstrated on a heterodimeric protein complex in translesion sysnthesis (TLS), a DNA damage tolerance pathway. The TLS machinery consists of several translesion DNA polymerases that are recruited to the stalled replication fork in response to monoubiquitinated proliferating cell nuclear antigen (PCNA) in order to bypass DNA lesions encountered during genomic replication. The recruitment and assembly of translesion machinery is heavily dependent on ubiquitin-binding domains, including ubiquitin-binding motifs (UBMs) and ubiquitin-binding zinc fingers (UBZs) that are found in translesion DNA polymerases. Two conserved ubiquitin-binding motifs (UBM1 and UBM2) are found in the Y-family polymerase (Pol) ι, both of which contribute to ubiquitin-mediated accumulation of Pol ι during TLS. Although the Pol ι UBM2-ubiquitin complex has been previous reported by our lab and others, the Pol ι UBM1-ubiquitin complex has remained a challenge due to significant signal overlap in conventional 3-D NOE spectroscopy. In order to determine the molecular basis for ubiquitin recognition of Pol ι, we solved the structures of human Pol ι UBM1 and its complex with ubiquitin by 4-D fast NMR, revealing a signature helix-turn-helix motif that recognizes ubiquitin through an unconventional surface centered at L8 of ubiquitin. Importantly, the use of 4-D omit NOE spectroscopy unambiguously

revealed an augmented ubiquitin binding interface that encompasses the C-terminal tail of UBM1.

4-D omit NOE spectroscopy was also used to study the Fanconi anemia associated protein 20 (FAAP20)-ubiquitin complex within the Fanconi Anemia (FA) complexes required for efficient repair of DNA interstrand crosslinks (ICLs), a process that is mediated by the ubiquitin-binding zinc finger (UBZ) domain of FAAP20. Unexpectedly, we show that the FAAP20-ubiquitin interaction extends beyond the compact UBZ module and is accompanied by transforming the disordered C-terminal tail of FAAP20 into a rigid β-loop, with the invariant C-terminal tryptophan (W180 of human FAAP20) emanating toward I44 of ubiquitin for enhanced binding. Accordingly, alanine substitution of the absolutely conserved C-terminal tryptophan residue of FAAP20 abolishes ubiquitin binding and impairs FA core complex-mediated ICL repair *in vivo*.

Reliable detection and unambiguous assignment of intermolecular NOEs is essential to NMR-based structure determination of protein complexes. The development of 4-D omit NOE spectroscopy in this thesis overcomes many limitations of conventional 3-D half-filtered experiments to allow for reliable detection and unambiguous assignment of intermolecular NOEs of heterodimeric complexes and homooligomeric complexes. These advantages render such a strategy particularly attractive for structural studies of protein complexes by biomolecular NMR.

Dedication

This dissertation is dedicated to my dearest husband Eric, my cutest son Aiden, my most wonderful Mom and Dad, Jing Su and Hanjun Wang.

Contents

List of Tables

List of Figures

Acknowledgements

At 8:00 pm on August 9th, 2008, I arrived at Dulles airport in Washington DC. That was the first moment I stepped onto U.S. soil. At that time, I couldn't have predicted what would happen during the more than six years I spent at Duke, but just wished I could always do my best. Now, more than six years have passed, during which time I met my husband, I got engaged and married, I had my son, all events that have dramatically changed my life. Most importantly, I am about to finish my PhD studies at Duke! This could not have been achieved without the support of so many people, both on a professional and personal basis.

The first person I would like to thank is my advisor Pei Zhou. He set a wonderful example as a researcher for us to follow with his intelligence, work ethic, passion for science, and careful attention to every detail of his research. I will never forget the many days and nights we spent in the Duke NMR center, working on all the details of pulse sequences as I joked: "NMR center is like a black hole, swallowing time without us noticing." I will also never forget our countless discussions about research, with Pei inspiring and encouraging me to face all the challenges and difficulties in research positively since I always liked to look at the "what if" side of things. Moreover, I really appreciate his understanding and support during the last period of my graduate studies, in which I struggled to balance my high-maintenance son and my research.

Also, I would like to thank all the past and current lab members in Pei's lab. I did my rotation with Jon Werner-Allen at a time when I knew nothing about structural study as I came to Duke Biochemistry with an immunology background. From him, I learned many good habits of doing science, such as being organized and paying attention to every detail. I thank him for giving me an excellent beginning in doing research. The UBM1-ubiquitin project was handed to me from Martha Bomar. Her previous studies of UBM2-ubiquitin gave me good examples to work from. Brian Coggins is the pioneer who developed 4-D fast NMR, so this work benefits greatly from the advances of the software that he developed. I would also like to thank Chuljin Lee for giving me countless tips and ideas for my projects. Jiangxin Liu and I sat back to back for several years and she has provided me with tons of support both experimentally and personally. Qinglin Wu was a tremendous help when I was having trouble with the NMR spectrometer or with data processing. Joshua Rose is such a considerate person and brought a lot of happiness to our lab. Jessica Wojtaszek and I worked on the UBZ project together, and her great ideas helped develop and further our project. I would also like to thank the other people in our lab including Jinshi Zhao, Hayley Young, Ali Masoudi, Jae Cho, Javaria Najeeb and Jenny Xue. Together, we are an incredibly happy and cooperative research family.

I would also like to thank our collaborator, Dr. Hyungjin Kim and Dr. Alan D. D'Andrea in the Dana-Farber Cancer Institute, for conducting the *in vivo* MMC

sensitivity experiments for the UBZ project. I also appreciate the Duke NMR center, Tony Ribeiro, Ron Venters and Don Mikas for their technical and instrumental support of my research projects. I would also like to thank the Oas lab for using their CD instrument, Harvey Sage at Duke University for assistance with the UBM1 analytical ultracentrifugation data, Dr. George Dubay for helping with mass spectrometry and Cheng-Yu Chen in Bruce Donald's lab for helping with PRE and RDC sample preparation.

Also, I would especially like to thank my committee members, Dr. Leonard Spicer, Dr. Seok-Yong Lee and Dr. Tso-Pang Yao for providing innumerable ideas and suggestions for my research and all kinds of support.

At Duke, I also met many good friends, some of whom came to Duke with me on the same flight. We enjoyed each other's company and the beautiful life at Duke while also supporting each other during the doldrums of graduate studies. Now, almost everyone has moved on to new lives in different places, but the beautiful memories will always stay with me.

Last but not least, I like to thank my great husband, Eric Jiang. I am so lucky to have meet him and to have him for my whole life. During the more than six years that we have spent together so far, we have had what feels like millions of happy moments and gone through difficulties together. He is always incredibly supportive and cheers me up any time I am down. My wonderful son has brought me tons of happiness and

sleepless nights. I would also like to give special thanks to my great parents, who always give me all they can to support my choices, love me unconditionally, never give me any pressure and always provide the freedom to pursue my passions. I love you all!

1. Introduction to structural study of protein complexes

As the functional units within a living organism, proteins rarely act alone at both cellular and systemic levels. Instead, protein complexes carry out diverse essential molecular processes within a cell such as gene transcription, translation, signal transduction, cell growth, differentiation and immune response. Abnormal protein-protein interactions that affect the formation and the function of normal protein complexes can lead to disease. In addition, proteins in eukaryotic cells are subject to various post-translational modifications such as phosphorylation, methylation, glycosylation and ubiquitination, which may further serve as signaling posts to recruit additional binding partners or to form higher order complexes, greatly enriching the diversity and functionality of protein complexes. Due to their essential roles in cellular function and disease, structural analyses of these protein complexes either by X-ray crystallography or NMR have not only shed important insights into the fundamental biology at the molecular level, but also provided important targets for disease intervention and developing novel therapeutics.

1.1 Heterodimeric protein complexes

A heterodimeric protein complex is the most common and simplest heteromultimeric protein complex. One interesting phenomenon of heterodimeric protein complexes is that the two proteins that need to bind to each other temporarily can be linked by interactions between a defined subdomain in one protein and a

chemical group that is transiently attached to the other. A classic example of this process is provided by protein domains that bind to phosphorylated amino acids, exemplified by the Src homology 2 (SH2) domain binding to tyrosine phosphorylated proteins (1). Ubiquitin is another type of post-translation modifier that labels proteins in a highly controlled matter (2). Ubiquitin recognition by ubiquitin-binding domains regulates a variety of biological events in the cell, including protein degradation, DNA repair, stress response and other events.

1.1.1 Ubiquitination

Ubiquitin, discovered in 1975, is a small protein with a highly-conserved 76 amino acid sequence among eukaryotic organisms (3); yeast and human ubiquitin only differ by three residues. It is named for its extraordinarily ubiquitous distribution in cells from yeast to humans. Early interest in ubiquitination centered on the role of polyubiquitin chains in targeting proteins for degradation by the 26 S proteasome (4, 5). Accumulating studies have shown that ubiquitination regulates a much wider array of cell processes, including endocytosis, vesicular trafficking (6-8), cell-cycle control, stress response, DNA repair (9), signaling (10, 11), transcription and gene silencing.

The three-dimensional structure of ubiquitin was first determined by X-ray crystallography in 1985 to reveal its α/β roll topology, now named as "ubiquitin fold" or "ubiquitin superfold"(12, 13). It contains a five-stranded β-sheet with a 3.5 turns α-helix on top, a short 3_{10}-helix and an exposed C-terminal tail with the last two glycine residues

being flexible. Ubiquitin is a remarkably stable protein, and is resistant to changes in pH and temperature.

Ubiquitination changes the molecular landscape of a protein, and can therefore influence the interactions of a protein with other proteins and, potentially, the three-dimensional structure of a protein. In the "protein ubiquitination" process, the C-terminus of free ubiquitin is activated by a ubiquitin activating enzyme (E1), passed on to a ubiquitin-conjugating enzyme (E2), and finally transferred onto an amino group with the aid of a ubiquitin ligase (E3) (14, 15). This results in the addition of a ubiquitin moiety to either the ε-amino group of a lysine residue or the extreme N-terminus of a polypeptide. Besides the modification of heterologous proteins, the targeted lysine can also reside within a second ubiquitin molecule, resulting in the formation of poly-ubiquitin chains. There are very few E1 enzymes (only one in yeast and about ten in human), more E2 enzymes (11 in yeast and about 100 in human) and many E3 ligases (54 in yeast and about 1000 in human), which determines the specificity of both the substrate and linkage type (16). Deubiquitylases (DUBs) reverse this modification by hydrolyzing the linkage between ubiquitin moieties or between ubiquitin and the substrate, rendering it a truly flexible signaling tag to regulate different cellular events (17, 18).

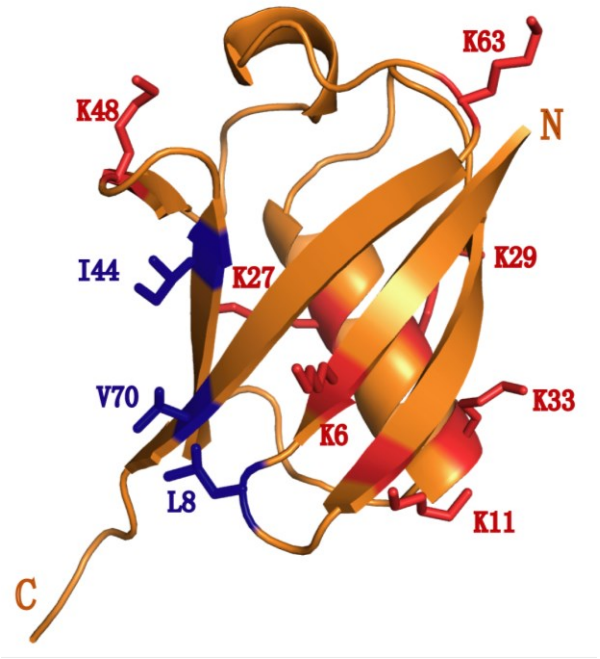

Figure 1: Structure of ubiquitin

The structure of ubiquitin is shown in cartoon mode, colored in orange. The key residues, L8, I44 and V70, which forms the canonical hydrophobic patch for interacting with ubiquitin-binding domains (UBDs), are shown in the stick mode and colored in blue. Seven lysines from which can form polyubiuqitin chains are colored in red. PDB ID: 1UBQ

Ubiquitin contains seven conserved lysine residues (K6, K11, K27, K29, K33, K48, and K63), that are highlighted in Figure 1. Therefore, there are eight homotypic chains (a chain that comprises only one linkage type) and numerous branched or non-branched heterotypic chains (a chain with two or more linkages). (Figure 2) These various types of ubiquitin chains further highlight the diversity of ubiquitination in different pathways.

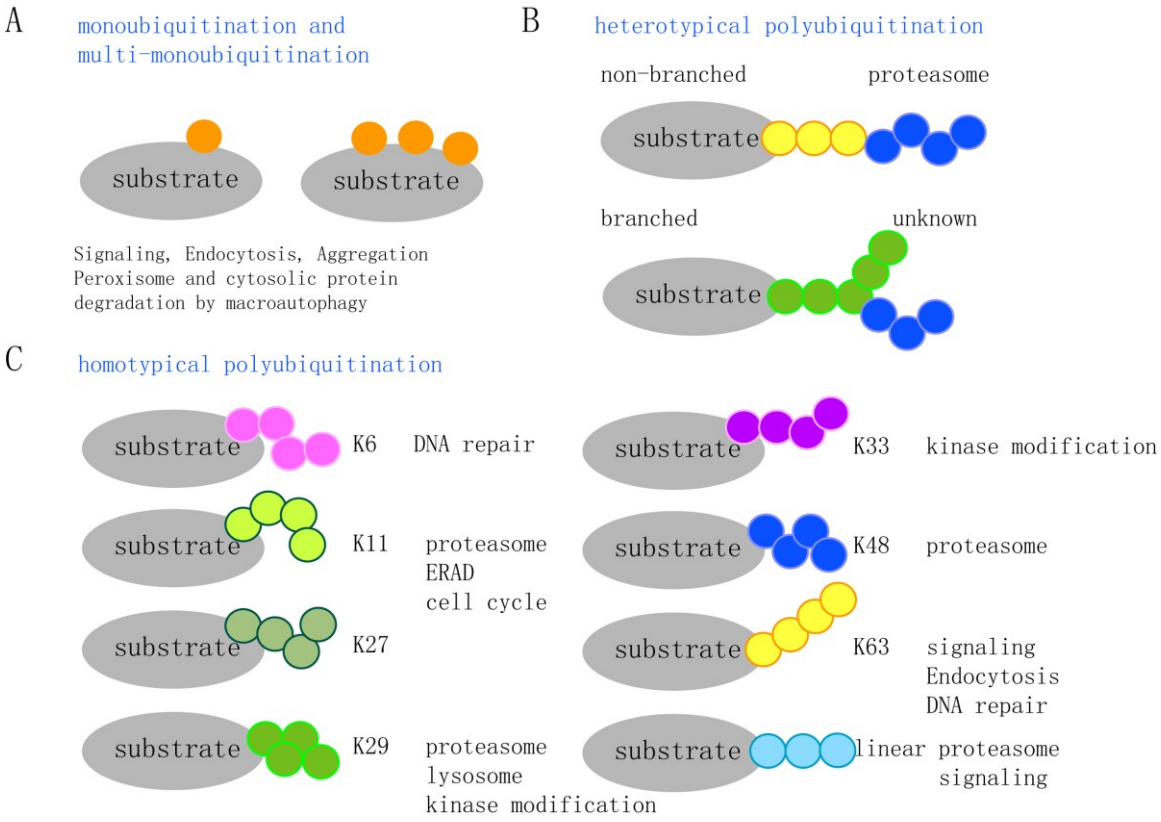

Figure 2: Diversity of ubiquitination

The ubiquitin molecule can be attached to a single site or multiple sites (A) on a substrate to yield mono- and multi-ubiquitination, respectively. (C) The ubiquitin sequence contains seven lysine residues that can support the assembly of polyubiquitin of different chain topologies, forming eight types of homotypical chains. Heterotypical polyubiquitination (B) occurs when a ubiquitin chain has alternating linkage types (non-branched linkages) or when a single ubiquitin is extended at two or more lysine residues (branched linkages). Adapted from (19).

Different ubiquitin linkages have been shown to be involved in different pathways (20, 21). For example, monoubiquitination plays important roles in signaling pathways involving membrane protein trafficking and endocytosis (22, 23). Of the eight distinct linkage types, K48-linked and K63-linked polyubiquitin chains are the most thoroughly characterized, with K48-linked polyubiquitin acting as a targeting signal for

proteasomal degradation (24). Ubiquitination can also promote non-degradative outcomes. First identified in 1995, K63-linked ubiquitin chains are involved in cell signaling (e.g. NF-κB pathway) and are essential for the DNA damage response (25). Besides K48 and K63 linked chains, other less abundant types of ubiquitin chains may have their dedicated physiological role upon activation of their cognate assembly machinery under certain conditions (26). For example, the formation of linear ubiquitin chains (head to tail) by LUBAC (linear chain ubiquitin chain assembly complex) was recently discovered and is essential for the TNFα-stimulated NF-κB pathway (27-29). The abundance of the K11 linkage type, which is assembled by the APC/C (anaphase promoting complex, also known as the cyclosome) during mitosis, dramatically rises when cells exit mitosis. At this point, the APC/C becomes activated and acts as a proteasomal degradation signal in cell cycle regulation (30). Recent studies showed K27-linked polyubiquitin plays dedicated roles in mitochondrial maintenance or mitophagy (31). Structurally, both K63- and linearly linked diubiquitin have similar extended conformations, with little or no contact between the ubiquitin moieties, whereas K48 chains allow for the ubiquitin moieties to pack against each other (32). The formation of these various ubiquitin chains further expands the versatility of ubiquitin as a signal.

1.1.2 Ubiquitin-binding domains (UBDs) and ubiquitin recognition

Ubiquitin and the distinct conformations of the ubiquitin chains can serve as signaling components that trigger molecular events in cells. It does this by operating as a

reversible and highly regulatory signal for interacting with ubiquitin-binding domains (UBDs) in cellular proteins. Many molecular details of signal transmission from ubiquitinated proteins (substrates that are modified following various cellular stimuli) to effector proteins (ubiquitin receptors containing one or more UBDs) have been elucidated in the past decade (16, 32-34).

Structures of ubiquitin bound to different UBDs indicate that ubiquitin adopts an overall similar conformation, with only slightly different structural changes of the loop connecting the secondary structural elements. An extensive set of residual dipolar couplings (RDCs) measured by NMR for analyzing the dynamic behavior of ubiquitin on the ps to μs time scale in solution (35, 36) shows the bound ubiquitin structures correspond to an ensemble of structures sampled by free ubiquitin. This indicates that the ubiquitin inherently exhibits some degree of structural conformational heterogeneity, which contributes to the adaptive interface of ubiquitin in recognizing different UBDs.

UBDs are usually very small (20-150 amino acids) protein modules that are diverse both in structure and in the mode of ubiquitin recognition. The binding affinity between ubiquitin and UBDs are typically in the range of μM to mM. So far, more than twenty different families of UBDs have been identified, and the number is still growing (TABLE 1). Many of the families have at least one structural representative. Most of the

UBD-ubiquitin structures are UBD complexes with monoubiquitin, though structures of UBDs bound to linear, K48 or K63 linked diubiquitin are beginning to be elucidated (34).

Even though the sizes, topologies, and structures of these UBDs vary, most of these domains interact with the solvent-exposed β-sheet of ubiquitin, which contains a hydrophobic patch consisting of residues L8, I44, and V70 (Figure 1) and is surrounded by R42, K48, H68, and R72 (37, 38). Although ubiquitin has this canonical hydrophobic patch for protein-protein interaction, ubiquitin receptors utilize a variety of structural folds and unique binding modules for recognition of ubiquitin. The ubiquitin binding modules can generally be divided into four categories based on the fold: 1) α-helical structures, 2) zinc fingers (ZnFs), 3) ubiquitin conjugating (UBC) domains present in E2 enzymes and 4) plekstrin homology (PH) folds.

The largest class of ubiquitin-binding domains adopts α-helical structures. The UIM (ubiquitin interacting motif), which is found in many trafficking proteins that recognize ubiquitinated cargos and the S5a subunit of the proteasome, consists of a single α-helix that binds to the hydrophobic pocket of ubiquitin formed by residues L8, I44, and V70 (39, 40) (Figure 3A). UIM has two variants, MUIM/IUIM (motif interacting with ubiquitin/inverted UIM) of Rabex-5 (the Rab5 exchange factor) (41, 42) and the double-sided UIM (DUIM) of Hrs (hepatocyte growth factor-regulated tyrosine kinase substrate) (43). MIU binds to ubiquitin in a similar manner to UIM with the α-helix in the inverted direction, whereas DUIM binds two ubiquitin molecules with equal affinity

using a short α-helix. In contrast, the UBA domain (ubiquitin-associated domain), the first UBD described, is a three-helix bundle in which two helices contact the canonical hydrophobic patch of ubiquitin (38, 44) (Figure 3F). The structure of the hHR23A UBA2 domain in complex with K48-linked diubiquitin shows that the UBA domain is sandwiched in between two ubiquitin moieties. It uses either α1-α3 or α2-α3 to interact with the I44-centered hydrophobic patch of ubiquitin (45) (Figure 4A). Similar cases have been found for CUE (coupling of ubiquitin conjugation to endoplasmic reticulum degradation domain) (46)and GAT (Gga and Tom1) domains (47, 48). The UBAN (ubiquitin binding in ABIN and NEMO proteins) domain of NEMO is a parallel coiled-coil dimer that binds to two linear ubiquitin chains (49, 50) (Figure 4C). Ubiquitin binding of the UBAN domain occurs to the canonical I44 surface of one ubiquitin moiety and with a surface adjacent to I44 on the second ubiquitin moiety. This structure highlights the fact that differences in the linker regions in polyubiquitin chains necessitate very specific interactions with UBDs. Another type of α-helical UBD is UBM (ubiquitin-binding motif), which was identified in the translesion synthesis (TLS) pathways (51). It contains two α-helices for interacting with ubiquitin.

The second category of UBDs is zinc fingers (not to be confused with ubiquitin-binding zinc finger or UBZ below), which does not bind to the canonical hydrophobic patch of ubiquitin (Figure 3C). These domains recognize monoubiquitin by binding to three different regions on its surface. The nuclear protein localization 4 ZnF (NZF)

domain (52), which serves as a ubiquitin-binding adaptor protein in the endoplasmic reticulum-associated degradation (ERAD) pathway, consists of four strands that form a central set of "knuckles" through the coordination of four cysteine ligands to a central zinc ion. It binds to the hydrophobic surface of the ubiquitin β-sheet centered on I44 and V70. By contrast, the A20-type ZnF domain of RABEX5 contacts a polar surface of ubiquitin that is centered on Asp58, whereas its C-terminal α-helical MIU/IUIM domain binds to the canonical I44 face (41, 42). A different binding surface is recognized by the DUB isopeptidase T (IsoT), the PAZ (polyubiquitin-associated zinc finger or the BUZ domain of Ubp-M, which consist of five β-strands and two α-helices, forming a deep pocket that binds to the C-terminal residues of ubiquitin (53, 54). UBZs (ubiquiting-binding zinc fingers) are another type of UBD that contain a $\beta\beta\alpha$ classical zinc finger fold to interact with ubiquitin (49) and will be discussed in detail later.

There are other types of UBDs with different fold and interaction modes with ubiquitin. The E2 ubiquitin-conjugating (Ubc) enzyme UbcH5c binds via its β-sheet with ubiquitin's canonical hydrophobic surface (55) (Figure 3D). The GLUE (GRAM-like ubiquitin binding in EAP45) domain present in the EAP45 (ELL-associated protein of 45 kDa) subunit of the complex ESCRT (endosomal sorting complexes required for transport) folds into a split PH domain using amino acids from a β-strand, loop, and α-helix and associates with the hydrophobic surface of ubiquitin (56) (Figure 3E). The proteasomal receptor Rpn13 (Regulatory Particle, Non-ATPase-like 13) contains a PRU

(PH receptor for ubiquitin) domain that adopts a PH domain structure and uses loops to bind to the β-strand surface of ubiquitin (57) (Figure 3B).

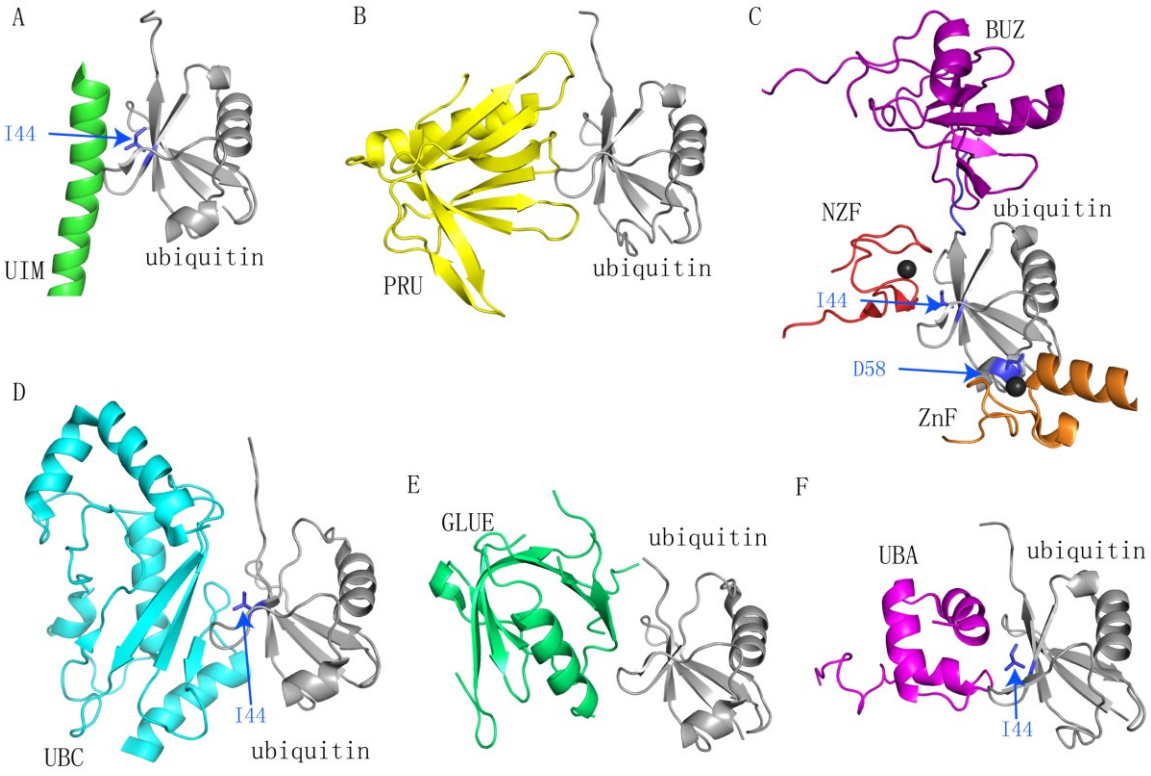

Figure 3: Structures of different types of UBDs in complex with monoubiquitin

Ubiquitin is shown in grey, with I44 or D58 or C-terminal tails colored in blue, which are critical residues in the interface. (A) UIM (green) in complex with ubiquitin, PDB ID: 1YX5; (B) PRU (yellow) in complex with ubiquitin, PDB ID: 2Z59; (C) different types of zinc finger domains interact with different interfaces of ubiquitin, showing A20 type-ZnF domain (orange) binding to D58 of ubiquitin, PDB ID: 2FIF, NZF domain (red) binding to I44 of ubiquitin, PDB ID: 1Q5W and BUZ domain (purple) binding to C-terminal tail of ubiquitin, PDB ID: 2G45; (D) UBC (cyan) in complex with ubiquitin, PDB ID: 2FUH; (E) GLUE domain (lime) in complex with ubiquitin, PDB ID: 2DX5; (F) UBA domain (magenta) in complex with ubiquitin, PDB ID: 2JY6

Table 1: The functional and structural diversity of ubiquitin-binding domains

UBD	Representing proteins	Functions	References
α-helical			
UIM	S5a/Rpn10, Vps27, STAM, epsins, Rap80(UIMC1)	Proteasomal degradation, endocytosis, MVB biogenesis, DNA repair	(40, 58-60)
MIU(IUIM)	Rabex-5	endocytosis	(41, 42)
DUIM	HRS	MVB biogenesis	(43)
UBA	Rad23/hHR23A, NBR1, Ede1, p62, Dsk2	Proteasome targeting, kinase regulation, autophagy, endocytosis	(44, 61, 62)
UBAN	NEMO, ABIN1-3, optinurin	NF-κB signaling, linear ubiquitination	(49, 50)
UBM	Pol ι, Rev1	DNA damage tolerance	(63, 64)
GAT	GGA, TOM1	MVB biogenesis	(47)
CUE	Vps9, TAB2, TAB3	Endocytosis, kinase regulation	(46, 65)
VHS	STAM, GGA3	MVB biogenesis	(66)
Zinc finger			
UBZ	Pol η, Pol κ, NEMO, Rad18, WRNIP, Tax1BP1	DNA damage tolerance, NF-κB signaling	(67-69)
NZF	Npl4, Vps36, TAB2, TAB3	ERAD, MVB biogenesis, kinase regulation	(52)
BUZ/ZnF-UBP/PAZ	IsoT, HDAC6, Ubp-M	Polyubiquitin chain disassembly, autophagy	(53, 54)
ZnF A20	Rabex-5 and A20	Endocytosis, kinase regulation	(41)
UBC-like			
UEV	Uev1, Mms2	DNA repair, MVB biogenesis and kinase regulation	(70, 71)
Ubc	UbcH5C	ubiquitination	(55)
PH domain			
PRU	Rpn13	Proteasome function	(57)
GLUE	Eap45 (VPS36)	MVB biogenesis	(56, 72)
Other			
SH3	Sla1, CIN85	Endocytosis	(73-76)
PFU	Doa1/Ufd3	ERAD	(77)
Jab1/MPN	Prp8	RNA splicing	(78)

Ubiquitin signaling appears to play a particularly prominent role in the DNA repair signaling pathways, including translesion synthesis (TLS) and repair of DNA interstrand crosslinks (ICLs). The ubiquitin-binding motif (UBM), which belongs to the α-helical family of UBDs and Ubiquitin-binding zinc finger (UBZ), which belongs to the zinc finger family of UBDs, are two typical kinds of UBDs that are dedicated to DNA damage response pathways (51, 79). The UBM domain, for example, has been shown to recognize ubiquitin through a non-canonical interface centered at L8 instead of I44. Previous studies from our lab and others have characterized the ubiquitin interaction of the C-terminal UBM (UBM2) of human translesion polymerase ι (64) and the UBZ of translesion polymerase η (68). However, the N-terminal UBM (UBM1) of polymerase ι displays significant sequence variation from UBM2, including the residues of the ubiquitin binding interface of UBM2, and hence requires further structural and biochemical investigations. Recent bioinformatics studies have also revealed distinct subfamilies of the ββα-UBZ fingers that are predicted to interact with ubiquitin using binding modes distinct from the well characterized UBZ3 subfamily represented by the UBZ of polymerase η, demanding further structural analysis of UBZs from other

subfamilies (79).

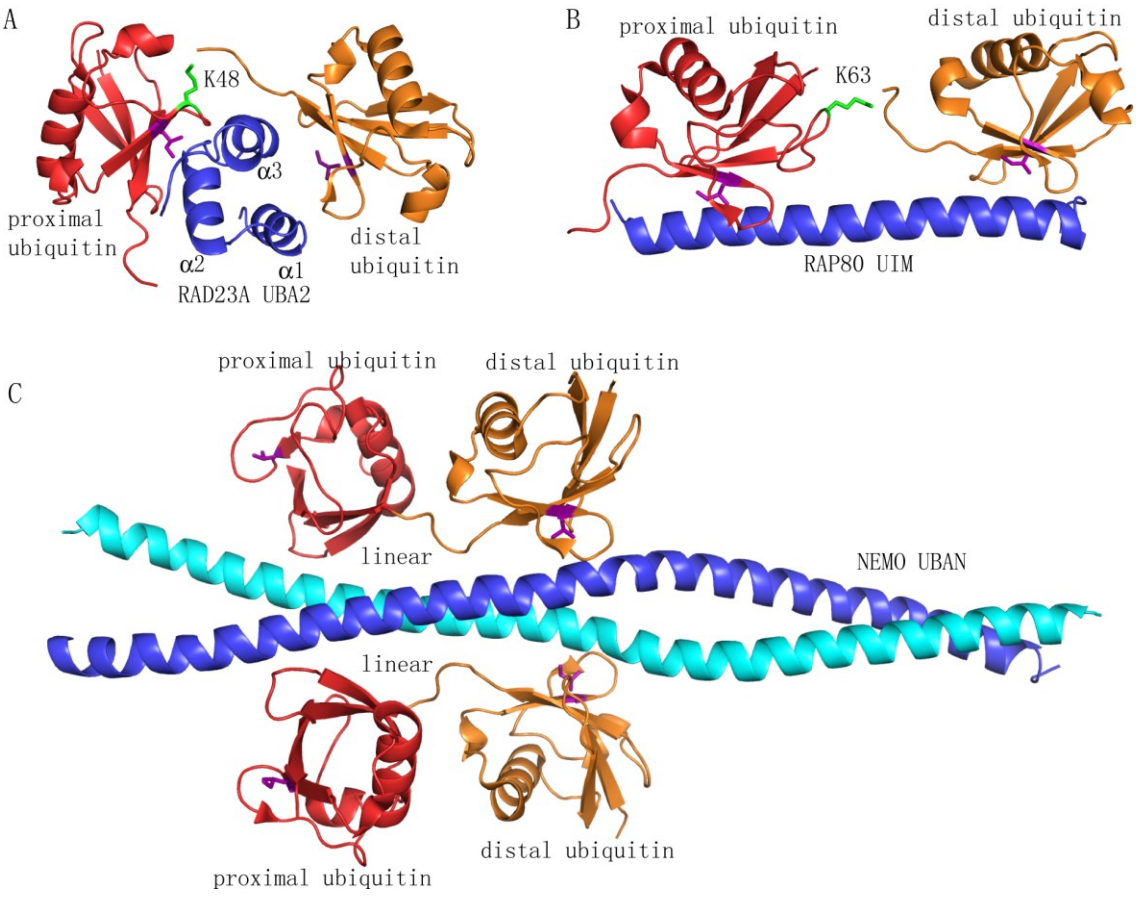

Figure 4: UBDs interacts with di-ubiquitin with different linkages

Multivalent interactions between ubiquitin and UBDs define chain specificity and increase affinity. Proximal ubiquitins are colored in red, distal ubiquitins are colored in orange, I44 are shown in stick mode and colored in purple. K48 or K63 are colored in green. (A)The UBA2 domain (blue) of RAD23A, is sandwiched between the two ubiquitin moieties of K48-linked diubiquitin, (PDB ID: 1ZO6) for expanding its binding surface beyond that possible for monoubiquitin. (B) Structure of K63-linked diubiquitin bound to RAP80, (PDB ID: 2RR9) showing its contiguous helix binds simultaneously to two ubiquitin moieties in the extended conformation for increasing its affinity. (C) The UBAN domain in NEMO forms a coiled coil, which binds two linear diubiquitins. (PDB ID: 2ZV0). Extensive contacts are formed with both ubiquitin moieties, thus conveying specificity for linear ubiquitin chains.

1.2 Homooligomeric protein complexes

Homooligomeric complexes are ones in which the interacting subunits are identical proteins. The majority of cellular proteins exist as symmetric oligomers with distinct biochemical and biophysical properties, which often provide the means for additional regulation of their function at the post-translational level (80-82). Homooligomers are found in enzymes, carrier proteins, scaffolding proteins and transcriptional regulatory factors. For example, *E. coli* diacylglycerol kinase (DAGK), a homotrimer, represents a family of integral membrane enzymes that is unrelated to all other phosphotransferases and is an important therapeutic target (83); Atm1-type (ATP)-binding cassette (ABC) transporters, a homodimer, plays an important role in the cellular detoxification processes (84, 85); D13, a homotrimer, is the scaffolding protein of vaccinia virus (86) and is the target of rifampicin. The famous tumor suppressor p53 also functions as homooligomer (87) through its tetramerization domain. Despite the essential biological function of homooligomer proteins, their structural analysis by solution NMR is particularly challenging due to the data interpretation ambiguity caused by the symmetry of the homooligomeric protein. Such challenges are highlighted by the structural discrepancy of the solution and crystal structures of DAGK, an integral homotrimeric membrane protein. The solution structure of DAGK, solved in DPC micelles, shows an atypical domain-swapping feature (88) (Figure 5A), whereas the

crystal structure (89) illustrates a different fold of the monomer itself, and the packing of the trimer does not involve domain-swapping (Figure 5B).

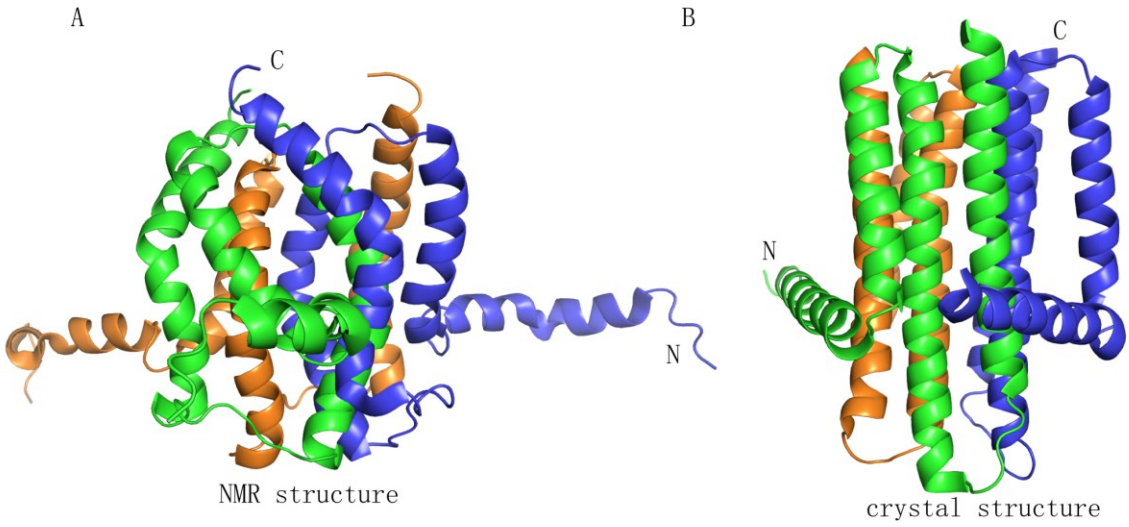

Figure 5: Comparison of NMR and crystal structure of DAGK

DAGK homotrimer shows completely different conformations as determined by NMR (A) (PDB ID: 2KDC) and crystal structure (B) (PDB ID: 4UP6).

The structural discrepancy of DAGK highlights the challenge of the structural analysis of protein complexes by solution NMR and encourages the development of new techniques to improve the reliability of the structural analysis of macromolecular assemblies. In this thesis, I outline the development of high-resolution 4-D omit NOE spectroscopy for structure determination of protein complexes. Such an approach has been applied to structural analyses of the homotrimeric foldon complex, the heterodimeric ubiquitin-Pol ι UBM1 complex and the ubiquitin-FAAP20 UBZ complex.

2. Development of 4-D omit spectroscopy for measurements of intermolecular NOEs

2.1 NMR studies of protein complexes

NMR remains as the only method for high-resolution structural elucidation of protein complexes in solution. It also readily reveals dynamic information that may not be accessible by crystallography. Recent pioneering work by Kay and co-workers have also demonstrated the feasibility of NMR to probe the structure and dynamics of minor populated high-energy states (90), opening up new frontiers for studies of transiently accessible conformational states of macromolecules and their complexes.

Traditional NMR structural analysis has relied on the measurements of Nuclear Overhauser Effect (NOE), the enhancement of one spin signal due to magnetization transfer from another dipolar coupled spin system through cross-relaxation. Commonly used NOE experiments include 3-D ^{15}N-NOESY that detects the NOEs towards amide protons and 3-D ^{13}C-NOESY that detects the NOEs towards ^{13}C-attached protons. Since the strengths of the NOE signals are inversely proportional to the distance (r^{-6}) of two interacting spins, they provide spatial restraints for NMR structural calculations. In addition to calculating an ensemble of NMR structures, quantitative analysis of NOEs can also reveal correlated motions within the structural ensemble (91).

Distance information provided by ^{1}H-^{1}H NOE crosspeaks with state-of-the-art spectrometers equipped with cold probes are typically restricted to 6 Å in fully protonated protein samples, though NOEs involving protons 10-15 Å apart have been

reported for deuterated proteins with selective methyl labeling. The NOE-derived distance information can be complemented with measurements of Paramagnetic Relaxation Enhancement (PRE). The large magnetic dipolar interaction of the unpaired electron from a paramagnetic atom with the neighboring NMR-active nucleus results in an increase of the relaxation rate of the nucleus (92). Similar to NOE, the PRE effect has the basic r^{-6} distance proportionality, but, because of the larger magnetic moment of the electron, the PRE effect is observable at longer distances, up to 25–35Å, distances depending upon the nature of the particular paramagnetic group (93, 94). Stable nitroxide radicals (eg. MTSL) (95) and metal chelators, such as EDTA-Mn^{2+}, that are characterized by an unpaired electron with an isotropic g-tensor, are especially useful for measuring the PRE effects.

Residual dipolar couplings (RDCs) are another set of commonly used NMR constraints for structural analysis. The physical basis of RDCs is the dipole-dipole (DD) interaction between two nuclear spins. In an isotropic environment, the dipolar coupling cancels out due to the free tumbling of the protein molecule. In a solution with partial alignment, an incomplete averaging of spatially anisotropic dipolar couplings results in the residual dipolar coupling (RDC) between two spins in a protein molecule, which is in the range that can be measured spectroscopically (typically 10-50 Hz). RDCs give the orientation information of each chemical bond relative to the external magnetic field B_0. Therefore, it provides global orientational constraints for NMR structure determination

or refinement. It is a useful tool that complements the NOE constraints and is particularly useful for measuring the bending of a long α-helix, for determining the orientation of two interacting protein domains, and for deciphering subunit orientations in a homooligomer system.

Dipolar coupling manifests itself as an additional coupling to the J-coupling. RDC measurements normally require two NMR samples prepared in parallel, one with the presence of an aligning medium and the other without, corresponding to anisotropic and isotropic conditions, respectively. Measurement in the isotropic medium yields J, while measurement in the anisotropic medium yields J +D (96), and the difference yields the dipolar coupling D. The commonly measured Ds are $D_{H\text{-}N}$, $D_{H_\alpha\text{-}C_\alpha}$ and $D_{C_\alpha\text{-}CO}$ (97, 98).

In order to measure RDC, a partially aligned media is needed for sample preparation. Common alignment media include liquid crystalline media(99), bicelles made of dimyristoylphosphatidylcholine (DMPC) and dihex-anoylphosphatidylcholine (DHPC) (100), filamentous phages Pf1 (101) and stretched or compressed polyacrylamide gel (102-104) and DNA based media (105). Addition of charged molecules into the alignment media has been shown to generate different alignment tensors for protein molecules (106). This is a particularly useful property as one type of alignment media is usually sufficient for structural refinement, whereas for *de novo* structure determination, any additional orthogonal alignment tensors from a new

alignment media add new information to better define bond vector orientation and dynamics.

In comparison with the long range information provided by NOEs, PREs and RDCs, measurements of vicinal scalar coupling between HN and Ha can provide dihedral angle constraints for determining the backbone dihedral angles ϕ and ψ (107). Similar information can also be extracted through analysis of chemical shifts of backbone atoms, which has been shown to correlate with the dihedral angles of the peptide bonds (108).

Despite the successful application of these aforementioned methods for structural analysis by NMR, there are limitations in particular with regard to the structural analysis of protein complexes. For the RDC methods, finding a suitable alignment media that is compatible with the target protein complex can be a challenge. Additionally, distinct sets of RDCs (e.g., both $^1D_{HN}$ and $^1D_{H_\alpha C_\alpha}$) measured with sufficient precision are required for *de novo* structure determination (109), which may be challenging for protein complexes beyond 15 kDa. Although measurements of PREs have provided long-range distance information, they are typically not as accurate as NOE-derived distances due to the intrinsic flexibility of either the paramagnetic tag itself or the flexible linker that it uses to attach to the protein. Such flexibility results in the time-averaged distances sampled over all possible conformations. Although PRE-derived long range constraints are most useful for defining the global fold of a target

protein, they are much less informative for defining interface interactions (110). Therefore, accurate measurements and analyses of intermolecular NOEs have become critical for structural analyses of protein complexes.

2.2 Current techniques for detection and assignment of intermolecular NOEs of protein complexes

Detection of intermolecular NOEs is traditionally achieved by half-filtered experiments (111, 112). The general strategy involves (1) isotopic labeling of a subset of the signals (usually a component of the protein complex) and (2) spectroscopic suppression of protons attached to ^{13}C or ^{15}N nuclei prior to NOE transfer followed by selective detection of ^{13}C or ^{15}N-attached protons for signal observation. Theoretically, this approach selectively detects dipolar interactions from protons in the unlabeled component to protons in the isotopically-labeled component within the same protein complex, thus ensuring all of the observed NOE crosspeaks are intermolecular. However, the large variation of $^{1}J_{HC}$ couplings from the methyl groups (~125 Hz) to aromatic groups (~200 Hz) makes it extremely difficult to achieve perfect suppression of isotope-attached proton signals. Thus, two solutions have been proposed to improve the robustness of half-filtered experiments.

The first approach is to incorporate adiabatic frequency swept carbon inversion pulses presented by Zwahlen et al. (113) The concept is based on the empirical relationship between $^{1}J_{HC}$ coupling constants with the isotropic ^{13}C chemical shifts, which is an approximately linear relationship for both proteins (Figure 6) and RNA. A filter

element can then be designed in which the conventional ^{13}C inversion pulse is replaced by a pulse whose frequency is swept adiabatically during the filter delay, such that inversion of ^{13}C spins occurs at a time that is a function of both offset and the value of $^{1}J_{HC}$ couplings. Therefore, the effective INEPT transfer times can match with the variation of $^{1}J_{HC}$ couplings.

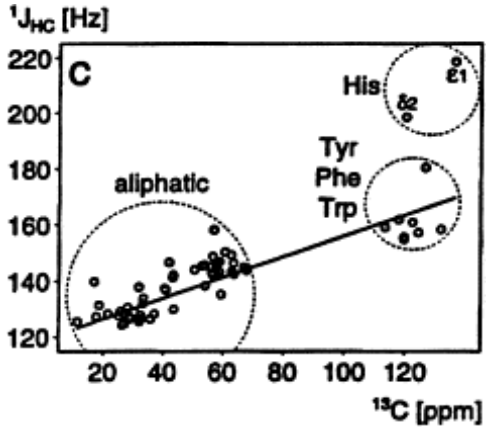

Figure 6: Correlation between $1J_{HC}$ and ^{13}C chemical shift in proteins.

The equation for the line derived from linear regression analysis is: $^{1}J_{HC}$ = (0.365 ±0.010 Hz/ppm) δC + 120.0 ± 0.5 Hz. Adapted from Zwahlen et al. (113)

The second approach is based on the isomorphism between broadband polarization transfer and composite pulse rotations (114). Composite pulses compensated for B_1 inhomogeneity are converted to sequence elements compensated for variation in $^{1}J_{HC}$ couplings. In addition, the entire filter period can be used for frequency labeling in a semiconstant time manner (115) to provide increased resolution in the indirect ^{1}H evolution dimension of 3-D half-filtered NOESY experiment. Also, since the filter efficiency does not depend on empirical correlations between ^{13}C chemical shifts

and $^{1}J_{CH}$ coupling, this approach will be useful for performing isotope-filtered experiments in weakly aligned systems for which RDC contributes to the apparent coupling interaction (116).

The performance of the isotope filter can be further enhanced by a combination of both approaches to generate joint composite-rotation adiabatic-sweep isotope filter sequences, or low pass J-filters (117). Combination of the two approaches yields much better filtration efficiency, compared with methods individually, particularly for ^{1}H spins with coupling constants that are near the extreme of the range of values or that do not satisfy empirical relationships between isotropic chemical shifts and scalar coupling constants.

Despite these elegant solutions, it is still challenging to achieve ideal suppression of intramolecular NOE crosspeaks. In publications that applied 3-D half-filtered NOESY experiments for detecting and assigning intermolecular NOEs for complex structure determination, the insufficiently suppressed intramolecular NOE peaks are usually labelled with a cross (Figure 7). This trend has remained constant in the NMR field for almost two decades. Even in very recent publications (118), cross-labelled artifactual peaks still exist in the spectrum, which complicates data analysis (Figure 7). Additionally, since the protein-protein interface often consists of similar functional groups (e.g., methyl-methyl packing) that share similar proton chemical shifts, these intermolecular NOE crosspeaks are often located very close to the diagonal positions.

Thus, incompletely suppressed diagonal signals, which are significantly stronger than the NOE crosspeaks, often cause assignment uncertainties. Furthermore, a lack of chemical shift dispersion in the proton dimensions suggests that even if the intramolecular NOEs can be suppressed with confidence, proton chemical shift degeneracy can still cause assignment ambiguities, precluding reliable assignment of all interfacial NOEs.

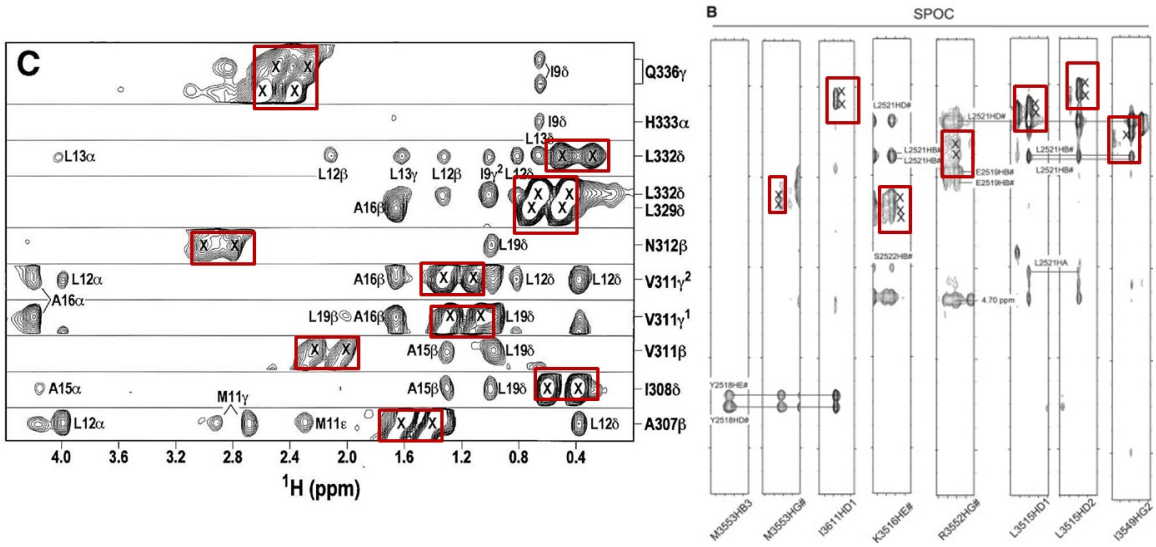

Figure 7: Examples of 3-D half-filtered NOE spectra

Figures adapted from current literature (left: (119); right (118)) showing 3-D half-filtered NOE spectra. Peaks labeled with crosses are boxed in red, which are explained as insufficient suppression of intramolecular NOEs.

The presence of significant levels of spectral artifacts in the 3D half-filtered NOE experiments coupled with assignment ambiguity may also lead to the determination of distorted or erroneously packed protein complexes. An example illustrating the limitation of 3-D half-filtered NOESY experiments is the solution structure of the

oligomerization domain of p53(120), which forms a dimer of a dimer. Each monomer consists of an α-helix and a β-strand. The β-strand from each monomer becomes antiparallel to form a dimer; while the dimer packs with another dimer through helical-packing. Compared with the 1.7Å crystal structure solved later on (121), the dimer-dimer interface in the solution structure is twisted (compare Figure 8A and B). After reanalyzing the NMR data, the authors found three mis-assigned intermolecular NOEs that should be spectrum artifacts and most importantly, three real intermolecular NOEs from the 3-D half-filtered NOESY experiment that were missed because they were treated as insufficient suppression of the intramolecular NOEs (Figure 8D) (122). One of the missed intermolecular NOEs, which was between two neighboring residues K351 and L350, was particularly strong. Due to the possible existence of residual intramolecular NOEs, it was conservatively assigned to be intramolecular. It turned out to be a key intermolecular NOE which correctly defined the dimer-dimer interface. After correction of the mistakes, the recalculated NMR structure overlaid very well with the crystal structure (compare Figure 8B and C).

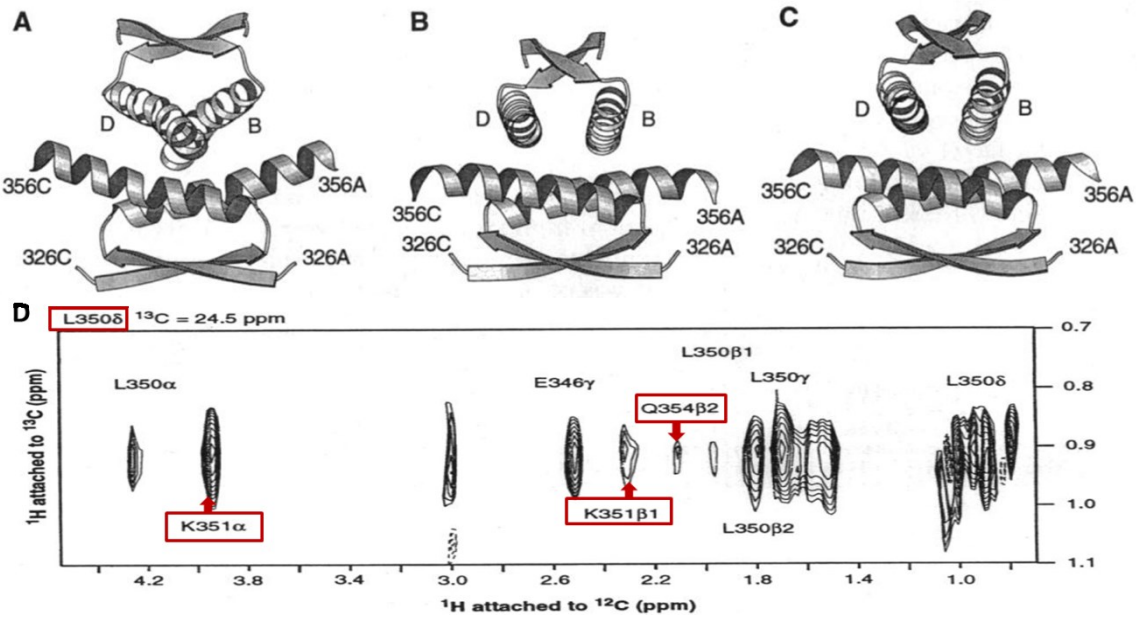

Figure 8: Structural study of oligomerization domain of p53

(A) Original solution NMR structure with tilted dimer-dimer interface; (B) crystal structure ; (C) corrected solution NMR structure; (D) Portion of the 3D ^{13}C-edited(F2)/^{12}C-filtered(F3) NOE spectrum (120 ms mixing time) of the oligomerization domain of p53 comprising a 1:1 mixture of unlabeled and ^{13}C/^{15}N-labeled polypeptide, specifically illustrating inter-subunit NOEs involving the methyl protons of Leu350. Real intermolecular NOEs peaks that were misassigned to be insufficient suppression are boxed in red. Adapted from Clore, M.G. et. al. (122).

These examples illustrate the limitations of current NMR methodology. In order to develop a robust method for structural analysis of protein complexes, we have developed 4-D omit NOE spectroscopy based on fast NMR technology and difference NOE spectroscopy.

2.3 Fast NMR

NMR experiments are traditionally limited by sensitivity, but advances in pulse sequence design (123) and instrumentation development, including higher field magnets

and cryogenically-cooled probes, have both boosted the sensitivity of NMR experiments (124). Because of these advancements, NMR has changed from a sensitivity-limited technique to a spectral resolution and dimensionality-limited technique. As an example, high-dimensional NMR experiments are extremely useful for better signal separation and more reliable data analysis. However, these high-dimensional NMR experiments are constrained by the long experimental measurement times. The development of fast NMR, which uses sparse sampling to dramatically reduce experimental measurement times, has becomes essential for the next generation of NMR experiments (125).

Conventional sampling of NMR experiments uses the Nyquist interval to create a regular grid of sampling points in the time domain (Figure 9A). It systemically increases the chemical shift evolution period of each indirectly-encoded nucleus by a fixed amount of time. The Nyquist interval in each dimension is determined by the spectral width, and represents the lowest sampling rate required to unambiguously define the frequencies of the signals recorded in that dimension. Therefore, conventional sampling prescribes the minimum pattern needed to uniquely determine the position of peaks in multidimensional NMR spectra. However, a fundamental limitation of conventional sampling is that the number of sampling points grows exponentially as the dimensionality of experiments increases. For example, a 3-D experiment collected with 64-point resolution in each indirect dimension requires 64^2 (4,096) sampling points, which typically translates into ~36 hours of acquisition time, whereas a 4-D experiment

collected with 64-point resolution in each indirect dimension requires 64^3 (262,144) sampling points, which requires ~194 days for data acquisition. Such an unrealistically long acquisition time hinders the application of high-resolution 4-D experiments.

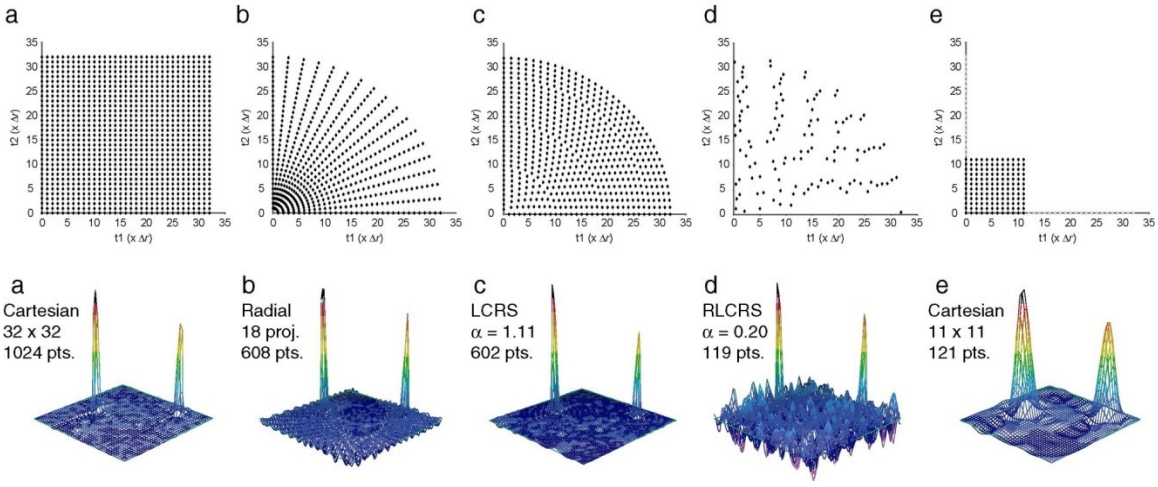

Figure 9: Fourier transform of sparse sampled data

The distribution of points that are being sampled in the time domain is shown for different sampling patterns in the upper panels. (A) is conventional Nyquist sampling, (B) is radial sampling, (C) is concentric ring sampling, (D) is random sampling and (E) low resolution Nyquist sampling. Axes represent indirect dimensions. Lower panels show the corresponding Fourier transform results for each sampling pattern. Adapted from Coggins et al. (126)

In order to speed up the acquisition of higher-dimensionality experiments, sparse sampling was developed, which uses a smaller number of points than required by conventional Nyquist sampling, thereby reducing the acquisition time. One of the first sparse sampling patterns introduced in the NMR field for 3-D experiments was radial sampling (Figure 9B). Such sampling places points in the time domain along radial spokes that meet at the origin (127). It is like sampling the time domain in the

polar coordinates instead of the Cartesian coordinates (128-130). The realization of radial sampling as a form of sampling in the polar coordinate has led to the use of discrete FT to process sparsely sampled NMR data and the interpretation of the associated "noises" being the aliasing artifacts (128, 131-133).

Besides radial sampling, there are a number of other sampling patterns including concentric ring or shell sampling (Figure 9C), random sampling (Figure 9D), and spiral sampling (134, 135). The distribution of sampling points on these patterns is usually weighted using an exponential or cosine function, biasing the measurement of low-resolution information found close to the origin of the time domain, in order to improve sensitivity and to impart a lineshape in the frequency domain.

These sparse sampling methods reduce the number of experimental measurements, which dramatically decrease the acquisition time of high-dimensionality spectra. The exact amount of time savings depends on the number of points sampled. The lower boundary of the sampling points depends on a variety of factors, but it should not be lower than the number of signals to detect. In reality, with ~1% of the sampling points, 4-D NOESY data can be successfully collected for protein around 30kDa within 4 days to identify a sufficient number of NOEs for structure determination.

Even though sparse sampling saves data acquisition time, this benefit comes at the cost of introducing aliasing artifacts. As described above, conventional sampling utilizes the minimum number of measurements required to precisely define the

frequencies of observed signals. Sparse sampling purposely uses numbers less than minimum which creates uncertainty in the frequency measurement of each signal, generating aliasing artifacts in the spectrum. There are several properties of aliasing artifacts: 1) the form of aliasing artifacts depends on the sampling pattern; 2) the intensity of aliasing artifacts depends on the extent of undersampling; 3) the level of artifacts generated by each signal is proportional to the signal's intensity. The third property is particularly problematic for spectra with high dynamic range, like NOESY spectra. The important NOE crosspeaks will be covered by the aliasing artifacts brought with strong diagonal signals.

To solve this problem, it is necessary to employ algorithms that remove aliasing artifacts as part of spectral processing following data collection. This can be realized because aliasing artifacts generated by a given peak can be predicted from the peak's intensity and spectral position and the sampling pattern. Following FT computation of the sparsely-sampled spectrum, the most intense peak is identified, and its artifact pattern is calculated based on the parameters described above. The peak and its artifacts are then subtracted from the spectrum, often revealing weaker peaks that were masked by the artifact noise. This process can be repeated iteratively until some stopping criteria, usually based on the residual noise level or the ability to confidently identify more peaks for processing, is reached. Uncorrupted peaks are then added back to yield a spectrum with reduced artifact levels (Figure 10).

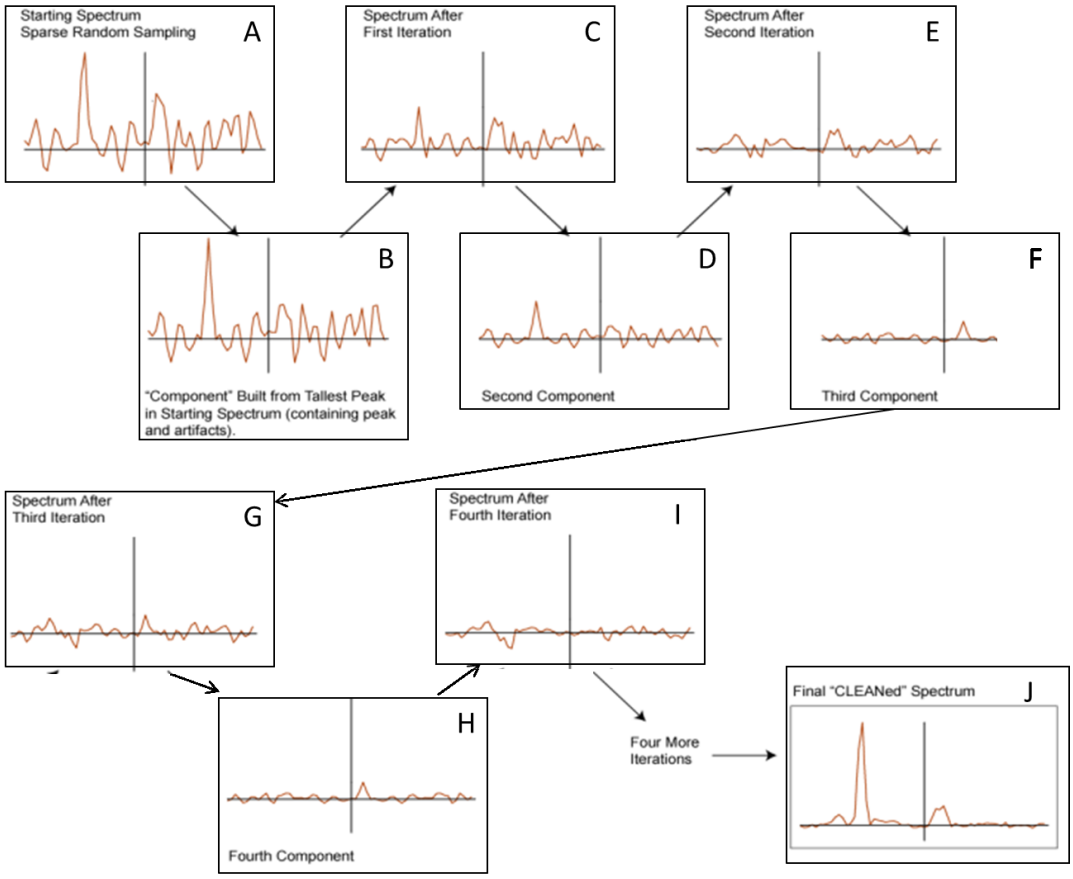

Figure 10: Removal of aliasing artifacts by CLEAN

Sparse sampling introduces aliasing artifacts after FT computation (A), from which the tallest peak is identified and its artifact pattern is calculated (B). The "Component" built from tallest peak in the starting spectrum (B) is subtracted from the starting spectrum (A) to generate (C), from which the second component is identified and built (D). This process can be repeated iteratively (E to I) to the basal residual noise level. Uncorrupted peaks are then added back to yield a final "CLEANed" spectrum with reduced artifact levels (J).

One algorithm following this concept is called CLEAN (134), which has been successfully applied to scalar coupling based backbone and sidechain resonance experiments, such as 4-D HCCH-TOCSY experiments (134). Such an approach can be

particularly beneficial in triple-resonance backbone experiments for sequential assignment of $^{15}N/^{13}C$ double labeled proteins (136). These triple resonance experiments utilize the relatively large magnetic couplings of the $1J_{NH}$, $1J_{CH}$, $1J_{CC}$, and $1J_{CN}$ to establish the scalar connectivity pathway between Hα, Cα, Cβ and CO atoms of neighboring residues. Conventional four pairs of triple-resonance experiments typically take over a week to collect, whereas implementation of the sampling technique (with 5% sampling density) followed by CLEAN reconstruction reduces the data acquisition time of the corresponding four pairs of 3-D triple resonance experiments—CA pair (HNCA and HN(CO)CA), CB pair (HN(CA)CB and HN(COCA)CB), CO pair (HNCO and HN(CA)CO) and HA pair (HA(CA)NH and HA(CACO)NH)—down to less than one day.

The ability to remove aliasing artifacts by CLEAN is not sufficient for use in the most challenging cases, such as NOESY spectra with diagonal signals. This is because the fractional subtraction of peaks by CLEAN can only allow each peak to be reduced down to the level of the apparent noise, at which point the residual signal can no longer be confidently distinguished. Thus, CLEAN leaves behind "stumps" for each peak at the level of apparent noise, along with the artifacts generated by the residual intensity. To solve this problem, a new algorithm called SCRUB (Scrupulous CLEANing to Remove Unwanted Baseline Artifacts) was developed (107) (Figure 11). SCRUB utilizes the basic CLEAN concept for removing artifacts, but is designed to process peaks in batches,

beginning with those that are strongest and well above the noise level, and applies special criteria to identify and include points adjacent to each peak's center, so that the full peak shape is subtracted rather than only the strongest point. In a simulated comparison, SCRUB reduces aliasing artifacts 250-fold further than CLEAN, making it well-suited for the reconstruction of sparsely sampled NOESY experiments. Indeed, exceptionally high-quality reconstruction of time-shared amide-methyl 4-D NOESY experiments has been achieved using SCRUB (107).

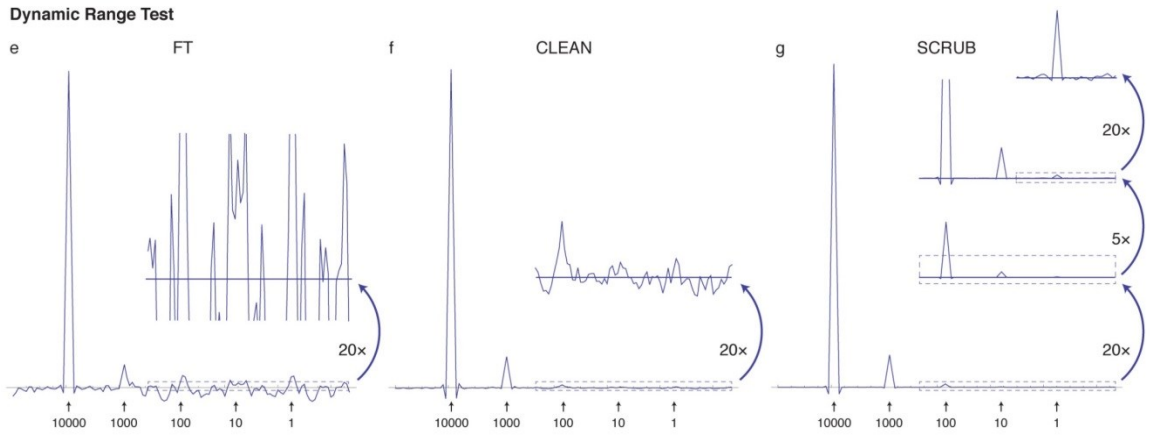

Figure 11: Comparison of DFT, CLEAN and SCRUB

Compared with DFT (left panel) and CLEAN (middle panel), SCRUB (right panel) has much higher dynamic range.

2.4 Development of 4-D omit spectra for detection and assignment of intermolecular NOEs of protein complexes

The improvement in sensitivity brought by the advancement of NMR instrumentation and the development of fast NMR technology with a high-quality reconstruction algorithm such as SCRUB has opened an exciting opportunity to develop

and apply novel experiments for studies of protein complexes. This is particularly important for NMR structural analysis of protein complexes, as the limited number of intermolecular NOEs that define the interface between the individual components in the macromolecular complexes require reliable detection and unambiguous assignment to ensure the proper assembly of the protein complex, which cannot be fulfilled with conventional 3-D half-filtered NOESY experiments as commented on previously.

In order to overcome the limitations of the conventional 3-D half-filtered NOESY experiment for detection of intermolecular NOEs, we developed a sparsely-sampled high-resolution 4-D ^{13}C HMQC-NOESY-HSQC experiment (Figure 12) to selectively detect NOEs from ^{13}C-attached proton to ^{13}C-attached proton.

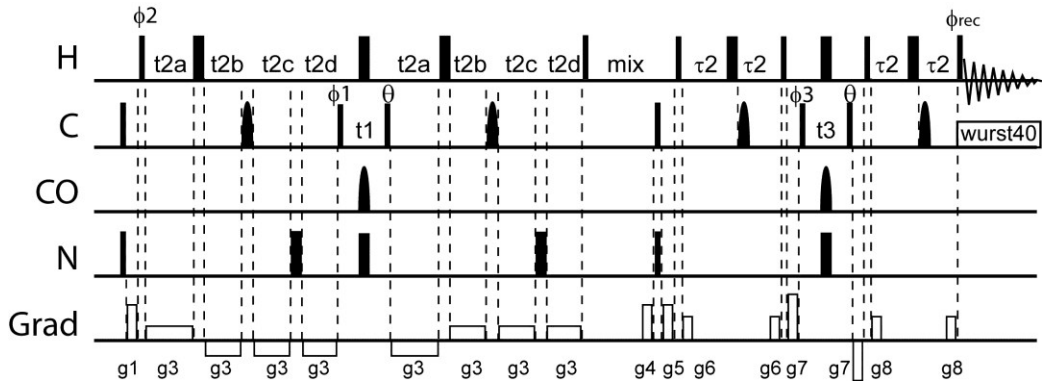

Figure 12: Pulse sequence for the 4-D 13C HMQC-NOESY-HSQC experiment.

Narrow and wide bars represent 90° and 180° pulses, respectively. All pulses are applied along the x-axis unless noted otherwise. Adiabatic broadband ^{13}C inversion pulses (500 μs) are used as previously described during INEPT transfer (113). Selective sinc-shaped ^{13}CO pulses are applied at 175 ppm with a null excitation at 56 ppm. An empirically optimized phase shift of θ=13° is applied to the 90° ^{13}C pulses after t1 and t3 to compensate Bloch-Siegert effects. A modified HMQC element was employed in the first two dimensions (F1-F2) to reduce signal relaxation based on the multi-quantum effect (137), whereas utilization of the HSQC element in the F3-F4 dimensions allows better water suppression using gradients. A semi-constant time evolution scheme is employed on both sides of the ^{13}C evolution periods in the HMQC element for recording ^{1}H chemical evolution and building the ^{1}H-^{13}C correlation. A shortened initial delay of tauCH1=1.6 ms is employed to reduce signal relaxation during the INEPT ^{1}H-^{13}C transfer step. The duration of the active $1J_{HC}$ coupling period is allowed to grow to tauCH3=2.1 ms when the proton evolution time permits. This modification enhances the sensitivity of the experiment for large proteins, but does not affect the quality of the signal selection. It should be noted that such an approach cannot be used in half-filtered experiments as in order to achieve ideal suppression of isotope-attached proton signals, a full delay is required for optimal buildup of $1J_{HC}$ correlations and subsequent suppression in the half-filtered experiment. The delays are t2a=tauCH1-κ*t2/4, t2b=(1-κ-γ)*t2/4, t2c=tauCH1+γ*t2/4, t2d=t2/4, mix=200 ms, τ2=1.8 ms. κ=tauCH1*4/t2max (for κ<1) or κ=1. γ=(tauCH3-tauCH1)*4/t2max (for γ < 1-κ) or γ=1-κ. ^{1}H and ^{13}C carrier frequencies are set to 4.8 ppm and 67 ppm, respectively. Carbon decoupling during acquisition is achieved by using a wurst40 sequence covering a bandwidth of 160 ppm (32.18 kHz on an 800 MHz spectrometer). Phase cycling is ϕ1=[x,-x], ϕ2=[x], ϕ3=[x,x,-x,-x], ϕrec=[x,-x,-x, x]. Gradient durations and field strengths are G1 = (1 ms, 15.72 G/cm), G3 = (3.88 G/cm), G4 = (1 ms, 19.80 G/cm), G5 = (1 ms, 18.58 G/cm), G6 = (0.5 ms, 12.86 G/cm), G7 = (1 ms, 20.21 G/cm), G8 = (0.5 ms, 12.45 G/cm). Quadrature detections in F1, F2, and F3 are achieved using the States-TPPI method by changing ϕ1, ϕ2, and ϕ3, respectively.

In addition to the ability to separate degenerate proton signals, a key feature of such a 4-D experiment is that it selectively detects NOEs between protons attached to ^{13}C nuclei. Hence, by subtracting NOE signals of subunit-selectively labeled protein complexes from the uniformly labeled protein complexes (Figure 13), one generates an "omit" spectrum–a concept analogous to the omit electron density map in crystallography (138) and the 2-D double difference NOE spectroscopy employing selective protein deuteration (139) – that only contains intermolecular NOE crosspeaks for unbiased data interpretation. Although such an approach is conceptually simple, it can only be achieved with a high-resolution 4-D NOESY experiment, but not with a 3-D NOESY-HSQC experiment, as the latter does not allow for selective detection of NOEs within isotopically-enriched subunits. Therefore, its application in biomolecular NMR has been prohibited by the lengthy data acquisition times of conventional high-resolution 4-D experiments (typically many weeks). The development of sparse sampling techniques (127, 140) and robust reconstruction algorithms overcomes this obstacle by making it practical to collect high-resolution 4-D NOESY spectra using only a fraction of the measurement time compared to the conventional approach. Although sparse sampling introduces aliasing artifacts proportional to the intensity and square root of the number of the NMR signals, these artifacts can be suppressed using iterative reconstruction algorithms such as SCRUB (141). As an added benefit, by subtracting NOE signals of individually labeled components from the uniformly labeled protein complex

in the time domain, one not only removes the very strong diagonal signals, but also drastically reduces the number of NOE signals for reconstruction, both of which further enhance the quality of the reconstructed omit spectrum containing intermolecular NOEs. In order to make sure all the peaks in the 4-D omit NOE spectrum are intermolecular NOEs, the peak intensity of all the diagonal peaks from the uniform-labeled and subunit-labeled spectra are measured so that scaling factors for subtraction can be calculated to make sure the intensity of all the intramolecular NOEs are either below noise level or negative. In most cases, this slight over-subtraction will not affect the intensity of the intermolecular NOEs except in situations in which there is an intramolecular NOE peak that completely or partially overlaps with the intermolecular NOE peak.

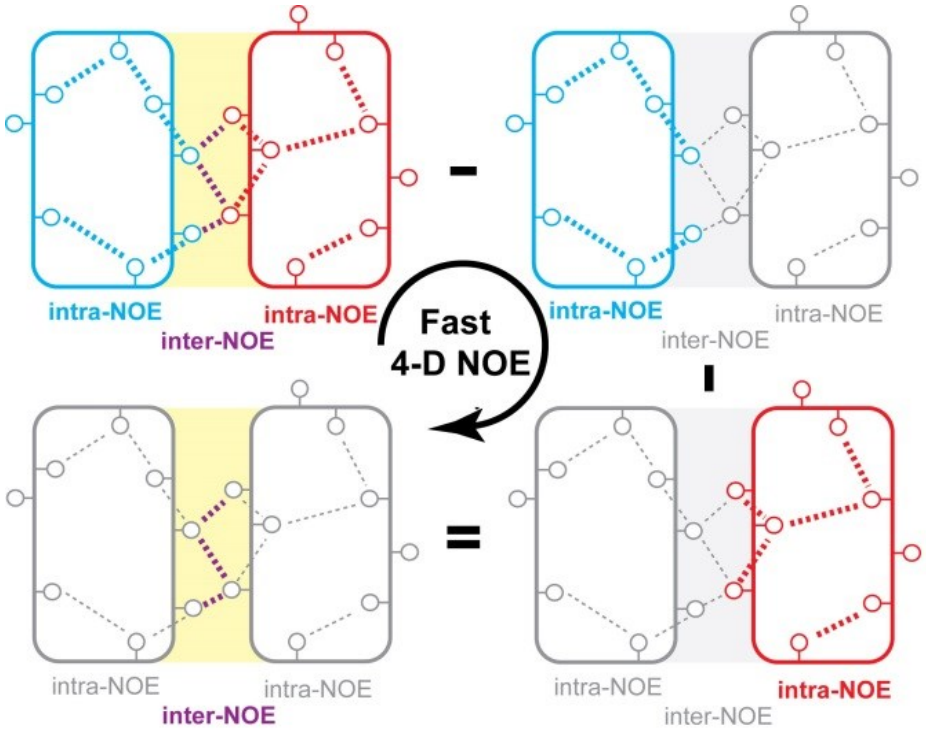

Figure 13: Concept of 4-D omit spectrum

Blue and red rectangles represent two interacting moieties in a heterodimeric protein complex. Yellow areas represent the interface. Circles represent protons, while dash lines represent the NOEs between protons. Colored portions denote isotopically-labeled and detected parts, while grey portions denote non-isotopically-labeled and undetected parts.

Compared with conventional 3-D ^{13}C half-filtered NOESY-HSQC experiments, there is little loss of sensitivity in a high-resolution 4-D ^{13}C HMQC-NOESY-HSQC experiment, as both experiments require similar durations of active J-coupling periods for building up H-C correlations, either for isotope signal suppression (as in half-filtered experiments) or for signal selection (as in the 4-D ^{13}C HMQC-NOESY-HSQC experiment). On the other hand, 4-D NOE spectroscopy offers much better signal separation than a 3-D half-filtered NOESY-HSQC experiment because of the extra C

dimension. Each peak is now identified from two pieces of chemical shift information: proton chemical shift and the directly-attached C chemical shift, which greatly reduces the assignment ambiguity. Additionally, because a 4-D ^{13}C HMQC-NOESY-HSQC experiment naturally selects *intramolecular* NOEs within the isotopically-labeled component of a protein complex, and because clean selection of isotope-attached proton signals is much easier to achieve than perfect suppression of the same signals, 4-D NOESY is much less affected by the variation of $^{1}J_{HC}$ couplings, thus ensuring the quality of intramolecular NOE spectra. Therefore, data analysis of 4-D omit spectra is much simpler and more reliable than analyzing 3-D half-filtered NOESY spectra.

As a proof-of-concept study, we applied this technique to the structural analysis of the T4 bacteriophage foldon trimeric domain.

2.5 Foldon

Bacteriophage T4 is a bacterial virus that consists of a prolate head, which contains DNA, and a six fold-symmetric tail, which is used for extruding viral DNA during infection. Fibritin is a 52 kDa protein product of the late gene wac (whisker antigen control) of phage T4. After the phage head is joined to a fiberless tail, six fibritin molecules attach to the "neck" of the virion through its N-terminal domain to form the collar structures ("whiskers"). Fibritin promotes the virus assembly, controls the retraction of the long tail fibers and is also the major component of the collar-whisker complex. Fibritin consists of an N-terminal anchor domain (residues 1-46), a large

central coiled-coil portion (residues 47-456), and a small C-terminal globular domain (residues 457-486) (142).

The thirty amino acid residue C-terminal domain is called foldon because of its essential function of fibritin trimerization and folding *in vivo* and *in vitro* (143, 144). The foldon domain is a trimeric β-sheet propeller consisting of monomeric β-hairpin segments. The trimer is stabilized by hydrophobic interactions involving Trp476 of each subunit, inter-monomer salt-bridges between Glu461 and Arg471, and inter-monomer backbone hydrogen bonds between Tyr469 and Arg471. It also contains an N-terminal hydrophobic stretch in the left-handed prolyproline II helix conformation between Pro460 and Pro463, which is connected to the β-hairpin (residues 468-479) and forms a hydrophobic cap of the hairpin on the N-terminal side. The hairpin ends in a 3_{10} helix at the C terminus with hemophilic interactions of several hydrophobic residues (Y458, I459, V470, L479, L483) between the monomers along the symmetry axis (145).

The foldon domain was proposed to be an evolutionarily optimized trimerization motif, since its only known function is to promote the folding of fibritin. At neutral pH, the foldon domain forms a stable trimer; however, when changing from neutral to acidic pH, the monomeric state appears in the NMR spectra of foldon, which is in slow exchange with the trimer state. The monomer state becomes increasingly populated between pH 4.3 and pH 1.6, and riches maximum population at pH values below pH 2 and concentrations of lower than ~200 μM (146).

Because of foldon's natural trimerization property, it has been engineered as a protein tag to promote the trimerization of other proteins that are fused to it (147, 148). And it is a model system for studying oligomer structure because of its small size, high stability and dynamic property between monomeric and trimeric states.

2.5.1 Detection and assignment of inter-monomer NOEs from 4-D omit NOE spectroscopy

The omit NOE strategy described above can be conveniently applied to homooligomeric complexes, such as the homotrimeric foldon complex. In this case, two sets of sparsely sampled 4-D ^{13}C HMQC-NOESY-HSQC spectra were collected for a 0.33 mM uniformly ^{13}C-labeled foldon sample and for a 1.32 mM sample containing 25% ^{13}C-labeled protein and 75% unlabeled (^{12}C) protein, respectively. Since intermolecular NOEs are statistically diluted in the second sample, signal subtraction of the second 4-D NOE dataset from the first dataset generates an omit spectrum containing intermolecular NOEs. The scaling factor was chosen so that all diagonal NOEs are slightly negative in the omit spectrum (e.g., the negative diagonal signal of A462 QB in Figure 14A), ensuring that the remaining positive NOE crosspeaks in the omit spectrum arise from genuine intermolecular NOEs (e.g., black colored crosspeaks between R471' and A462 in Figure 14A) and distinguishing these signals from the slightly negative intramolecular NOEs (e.g., intramolecular interactions between A462 and its neighbors; Figure 14). In total, 79 inter-monomer NOEs are identified from the 4-D omit spectrum.

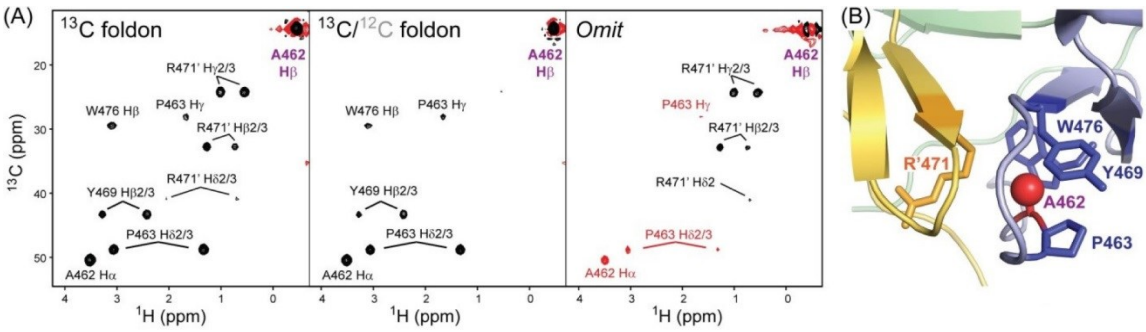

Figure 14: Omit spectrum of the foldon trimer

(A) F1-F2 slices of the corresponding 4-D spectra centered at 14.12 ppm in F3 and -0.49 ppm in F4, displaying NOEs to the A462 methyl group. 4-D ^{13}C HMQC-NOESY-HSQC spectra are collected for foldon with uniformly (left) or 25% (middle) ^{13}C-labeled samples. Subtraction of these two spectra generates an omit spectrum (right) containing positive intermolecular NOE and negative (red) diagonal signals and intramolecular NOEs. (B) Interface of two subunits of the foldon trimer, showing an inter-subunit interaction between R471 and A462 and intra-subunit interactions between W476, Y469, P463 and A462.

2.5.2 Automated structure calculation of foldon trimer

Since the structure of the foldon trimer is already known, we examined whether intermolecular NOEs detected in the 4-D omit NOE spectrum can be used for automated structural analysis of homooligomers. In order to achieve this, we first calculated the monomer structure of foldon using backbone resonances from 3-D sparse-sampled triple resonance experiments and sidechain resonances from 4-D sparse-sampled HC(co)NH-TOCSY experiment. 4-D sparsely-sampled ^{13}C-HMQC-NOESY-^{15}N-HSQC and ^{13}C-HMQC-NOESY-HSQC experiments were collected using 25% ^{15}N, ^{13}C-labeled samples to generate distance constraints for the monomer. CYANA (149) was used for automated structure calculation of the monomer.

Secondly, intermolecular NOEs between monomers were generated as described above and assigned automatically by CYANA (149) using a modified protocol to ensure that they are assigned as intermonomer NOE and are grouped according to C3 symmetry (see section 2.6.6 for detailed explanation).

Finally, intermolecular NOE constraints were combined with monomer structure constraints applied to each subunit to calculate the final foldon trimer structure.

Using intermolecular NOEs derived from the 4-D omit NOE spectrum, this strategy generates a highly-converged foldon structure with 0.6 Å backbone RMSD deviation from the corresponding crystal structure (PDB 1OX3) (Figure 15), demonstrating the benefit of 4-D omit NOEs in this automated calculation for a homooligomer.

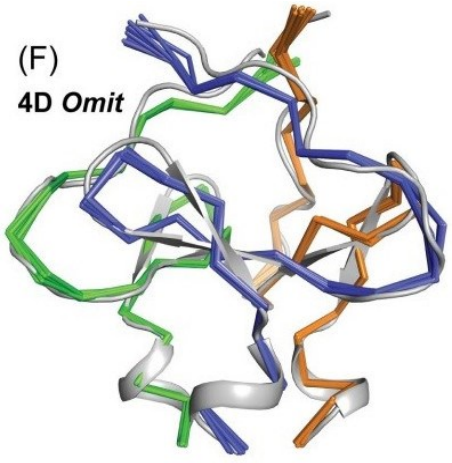

Figure 15: Automated structure calculation of foldon

The structural ensemble of trimeric foldon (subunits colored in blue, green and orange) calculated from the CYANA-assigned 4-D intermolecular NOE peak list from the omit spectrum superimposed with the crystal structure (grey; PDB ID: 1OX3).

2.5.3 Comparison of inter-monomer NOEs from 4-D omit NOE spectroscopy and 3-D half-filtered NOESY experiments for structure calculation

In order to compare our new 4-D omit NOE spectroscopy strategy with traditional 3-D half-filtered NOESY, we collected a 3-D half-filtered ^{13}C NOESY-HSQC experiment using the 25% ^{15}N, ^{13}C-labeled samples. NOE crosspeaks from this spectrum (Figure 16E) were picked and assigned following the same automated protocol as stated in Section 2.6.6 at different contour levels. However, due to assignment ambiguities, these crosspeaks cannot be analyzed by automated procedures in CYANA (149) to generate a properly converged fold, even when NOE crosspeaks are filtered at 30-fold and 90-fold above the spectral noise, respectively (Figure 16 C,D). This comparison demonstrated the advantage of our 4-D omit NOE spectroscopy in terms of homo-oligomer structure calculation.

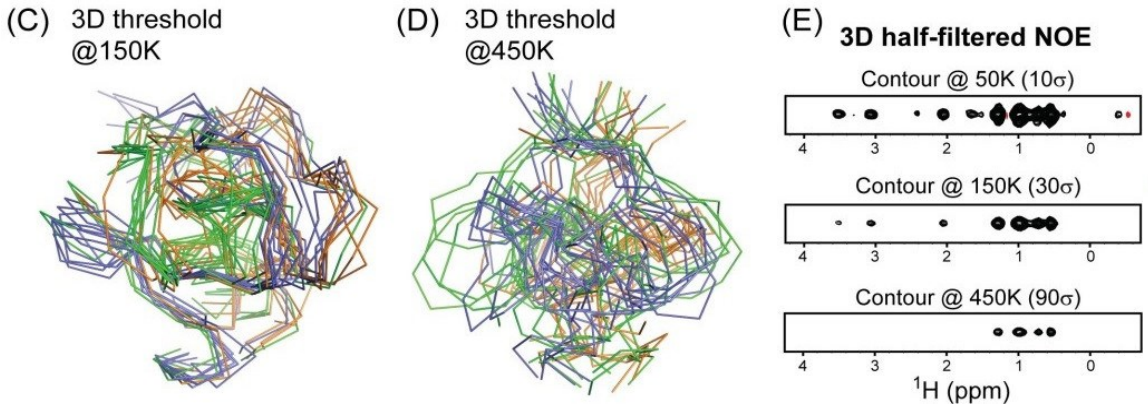

Figure 16: Automated structure calculation of foldon using 3-D half-filtered NOE

Panels C and D show the results of structure calculation using the 3-D half-filtered NOE peak list with intensity thresholds of 150K and 450K, respectively. E) A slice of the 3-D half-filtered NOE spectrum centered at the A462 methyl group, plotted at contour levels of 50K, 150K and 450K, respectively.

2.5.4 Pros and cons of applying 4-D omit spectra to homooligomer structure determination

The successful demonstration of the 4-D omit NOE approach to the automated structural analysis of the foldon trimer is encouraging, but not without caveats. For example, it is not possible to obtain *pure* intramolecular NOEs using a homooligomeric sample containing both isotopically-labeled and unlabeled components. Instead, one obtains a spectrum containing intramolecular NOEs and significantly weakened intermolecular NOEs due to *statistical* dilution of intermolecular dipolar interactions. After slight over-subtraction of subunit-labeled data from uniform-labeled data to make sure all the intramolecular NOEs are either invisible or negative, the intensity of the intermolecular NOEs in the 4-D omit spectrum will be reduced as well. Therefore, the 4-D omit spectrum in this case is less sensitive than the corresponding 3-D half-filtered NOE experiment in detecting intermolecular NOEs. However, as demonstrated with the foldon trimer, despite the superior sensitivity of the 3-D half-filtered NOE experiment (compare the omit spectrum in Figure 14 and the half-filtered spectrum plotted at different contour levels in Figure 16E), due to assignment ambiguities, these crosspeaks cannot be analyzed by automated procedures in CYANA (150) to generate a properly converged fold, even when NOE crosspeaks are filtered at 30-fold and 90-fold above the spectral noise, respectively (Figure 16CD). In contrast, the inclusion of an additional dimension in the 4-D omit spectrum largely eliminated the assignment ambiguity for the

automated NOE analysis protocol. Since the identity of the subunits contributing to intermolecular NOEs for a homooligomeric complex with C3 symmetry cannot be differentiated (e.g., interaction of A-B vs. A-C), these intermolecular NOEs are symmetrized in the order of A-B, B-C, and C-A, with the counter rotating pairs of A-C, B-A, C-B implemented as ambiguous constraints. When these constraints are used in the CYANA (150) calculation, they readily generate a highly-converged foldon structure with 0.6 Å backbone RMSD deviation from the corresponding crystal structure (PDB 1OX3), clearly demonstrating the advantage of 4-D omit NOE spectrum.

Although not demonstrated here, variation of the high-resolution 4-D NOE experiment developed for the omit NOE approach can also be used for direct detection of intermolecular NOEs between ^{15}N labeled monomer and ^{13}C labeled monomers, further improving the robustness of structural analysis of homooligomeric protein complexes by NMR.

The inability to detect pure intramolecular NOEs is a problem unique to homooligomeric protein complexes, which does not occur in heterodimeric or heterooligomeric protein complexes. This makes the omit NOE approach a much more attractive solution for studies of protein complexes consisting of distinct components. In the following two chapters, we apply the 4-D omit NOE experiments to studies of heterodimeric ubiquitin complexes with the Pol ι UBM1and the FAAP20 UBZ in DNA repair pathways.

2.6 Materials and methods

2.6.1 Molecular cloning

The DNA sequence of the T4-fibritin foldon domain (residue 457-483) was synthesized; the PCR-amplified DNA was double digested and ligated into a modified pTRC-HisB vector (Invitrogen) between the BamHI and EcoRI restriction sites. The final construct contains an N-terminal His_6 tag and thioredoxin tag followed by a thrombin-cleavage site. The insertion of the DNA fragment was confirmed by DNA sequencing.

2.6.2 Protein purification

The N-terminal His6 and thioredoxin tagged foldon construct were overexpressed in Escherichia coli BL21 (DE3*) cells (Invitrogen). Bacterial cells were cultured in M9 minimal media using ^{15}N-NH4Cl and ^{13}C-glucose as the sole nitrogen and carbon sources (Cambridge Isotope Laboratories), and induced by IPTG (1 mM IPTG at 37°C for 3 hr). The overexpressed proteins were purified by a Ni^{2+}-NTA column; then the N-terminal His6-thioredoxin tag of foldon were removed by thrombin cleavage, followed by passing through another Ni^{2+}-NTA column and benzamidine column to remove the His_6-thioredoxin tag and thrombin. Foldon protein was further purified by gel filtration column (Superdex 75, GE Healthcare).

2.6.3 NMR sample preparation and data collection

The uniform labeled NMR sample of foldon was prepared by buffer exchanging the protein after gel-filtration column into NMR buffer containing 10mM sodium

phosphate, 10%D2O or 100%D2O (pH=7.0) and concentrated to a final trimer concentration of 0.33mM.

The 25% labeled NMR sample of foldon was prepared by mixing ^{13}C, ^{15}N-labeled foldon with unlabeled foldon at a molar ration of 1:3 and incubating for 6 hr (pH=2.0). The foldon samples were then exchanged into the NMR buffer and concentrated to a final concentration of 0.33 mM for uniformly labeled foldon and 1.3 mM for the 25% labeled foldon, respectively.

NMR experiments were conducted at 25 °C using Agilent (Varian) INOVA 600 MHz or Agilent (Varian) INOVA 800 MHz spectrometers. NMR data were processed by NMRPIPE (151) and SCRUB (141). NMR spectra were analyzed with SPARKY (152) and XEASY (153). Backbone resonances were assigned by four pairs of sparse-sampled 3-D triple-resonance experiments, and the side-chain resonances were assigned using the 4-D sparsely-sampled HC(co)NH-TOCSY experiment.

2.6.4 Generation of 4-D omit spectra and identification of inter-monomer NOEs

4-D sparsely-sampled ^{13}C-HMQC-NOESY-^{15}N-HSQC experiment was recorded for 25% ^{15}N, ^{13}C-labeled sample, and identical 4-D sparsely-sampled ^{13}C-HMQC-NOESY-HSQC experiments were recorded for uniformly and 25% ^{15}N, ^{13}C-labeled samples. The 4-D sparsely-sampled ^{13}C-HMQC-NOESY-^{15}N-HSQC experiments used a mixing time of 150 ms, with spectral widths of 62ppm for the ^{13}C (t1 and t3) dimensions and 10.75 ppm for the ^{1}H (t2) dimension and t_{max} of 5.16ms for the C(t1 and t3) dimension and 11.16ms

for the h(t2) dimension. Each experiment was recorded for 86 hr. The omit spectrum was generated by subtracting scaled FIDs of 25%-labeled complexes from those of the uniformly-labeled complex and reconstructed by SCRUB (141). The scaling factor 1.7 was chosen so that all the diagonal NOEs are slightly negative in the omit spectrum. 4-D intermolecular NOEs were identified from the positive peaks in the omit spectrum.

2.6.5 3-D half-filtered NOESY and identification of inter-monomer NOEs

For comparison, 3-D half-filtered ^{13}C NOESY-HSQC experiment was recorded for the 25% ^{15}N, ^{13}C-labeled samples with a mixing time of 150 ms. The spectral widths were 12.75 ppm for the indirect ^{1}H (t1) dimension and 62 ppm for the ^{13}C (t2) dimension. It used a t_{max} of 7.85ms for the ^{1}H (t1) dimension and 4.84 ms for the ^{13}C (t2) dimension with a total acquisition time of 120 hr and a minimum required phase cycle of 16 scans. 3-D intermolecular NOEs were identified from this experiment at an intensity threshold level of 50k (a contour level of ~10-fold above the estimated spectral noise).

2.6.6 Automated structure calculation for homooligomer

Distance constraints for calculating the monomer structure were identified from 4-D sparsely-sampled NOESY experiments using a 25% ^{15}N, ^{13}C-labeled complex sample to ensure the detected NOEs are dominated by intramolecular signals, as in such a sample, the intermolecular NOEs are statistically diluted and significantly weakened. Using the 4-D NOE peaks derived from the 25% labeled sample and dihedral angle constraints derived from TALOS+ (108) analysis of chemical shift information; we were

able to calculate the monomer structure of the foldon trimer using the automated protocol of CYANA (149).

The default CYANA protocol for homodimeric structure calculation fails to work for protein complexes with C3 symmetry, as the default symmetrization method is incompatible with Cn symmetry. In order to overcome this issue and enable (semi-)automated structural analysis of the foldon trimer, we took the following approach. In the first step, the intermolecular NOE peaks extracted either from the 4-D omit spectrum or from the 3-D half-filtered ^{13}C NOESY-HSQC were used as input for the automated NOE analysis protocol (noeassign.cya) assuming a single chain foldon trimer. In this case, CYANA readily generates *unbiased* assignments, including intramolecular and intermolecular assignments, based purely on chemical shift information in the first cycle of structure calculation (cycle1.noa). In this step, two modifications were made to the default protocol (noeassign.cya): 1) Similar to the case of sparse methyl and amide NOEs for the global fold calculation, the weight of network anchoring was decreased in order to allow for the assignment of "lone" intermolecular NOEs; 2) Since we were interested in all intermolecular NOE assignments, the process of "constraint combination" was not used and the "cycle1-uncombined.upl" file, which contains *all* distance constraints, was used for generating intermolecular distance constraints. In the second step, the CYANA assignments were filtered to eliminate intramolecular assignment as the input peaklists from the 4-D omit spectrum or the 3-D half-filtered spectrum should only contain

intermolecular NOEs. In the third step, the intermolecular assignments were grouped based on symmetry. For homooligomeric protein complexes, there exists inherent subunit ambiguity. In the case of foldon with the C3 symmetry and subunits A, B, and C, an intermolecular NOE could occur between subunits A-B or A-C, but not necessarily both. As a result, six potential NOE assignments could be generated for an isolated signal, including three symmetry-related assignments of A-B, B-C, and C-A, and three alternative assignments of A-C, B-A, and C-B. These six assignments are grouped together as *two sets of three distance constraints*, with the alternative assignment implemented as ambiguous NOEs (i.e., A-B/A-C, B-C/B-A, C-A/A-B). In the final step, intermolecular NOE constraints were combined with monomer structure constraints applied to each subunit to calculate the final foldon trimer structure.

Using the intermolecular NOE peak list from the 4-D omit spectrum, such a protocol generated a highly converged foldon trimer that superimposed well with the reported crystal structure. However, when the intermolecular NOE peak list (generated by manual peak picking at a contour level of 50k) from the 3-D half-filtered experiment was used as input, the calculation failed to converge (data not shown). In order to eliminate the possibility that the 3-D half-filtered NOE spectrum could contain residual incompletely filtered intramolecular NOE signals, the 3-D peak list was filtered again at an intensity threshold of 150k (30-fold above spectral noise) and at 450k (90-fold above

the spectral noise). The filtered peak lists with different intensity thresholds were used as input to calculate the structure ensembles shown in Figure 16C and 16D.

3. Structural study of the human polymerase ι ubiquitin-binding domain 1 (UBM1) in complex with ubiquitin

3.1 Background

3.1.1 Ubiquitin recognition in translesion synthesis

3.1.1.1 Tranlesion synthesis

Translesion synthesis (TLS) is a type of DNA damage tolerance mechanism in which DNA damage can be bypassed by specialized DNA polymerases (translesion polymerases) during genomic replication. Due to the low replication fidelity of translesion polymerases, TLS is an error-prone process and it promotes cell survival at the cost of replication fidelity (154, 155). Most of the TLS polymerases belong to the recently classified Y-family of DNA polymerases (156, 157), with the exception of pol ζ (Rev3/Rev7), which is a B-family polymerase (158).

There are five subfamilies within the Y family: Rev1, UmuC, DinB/pol κ, pol η, and pol ι (157). Over 300 members have been identified to date within this family (159). *E. coli* has two TLS polymerases, Pol IV (DinB/pol κ) and Pol V (UmuC); *S. cerevisiae* also has two (Pol η and Rev1), while human has four Y-family polymerases: Pol η, Pol ι, Pol κ, and Rev1 (Figure 17). These polymerases differ in their preferred lesion substrates and catalytic efficiencies towards various lesions.

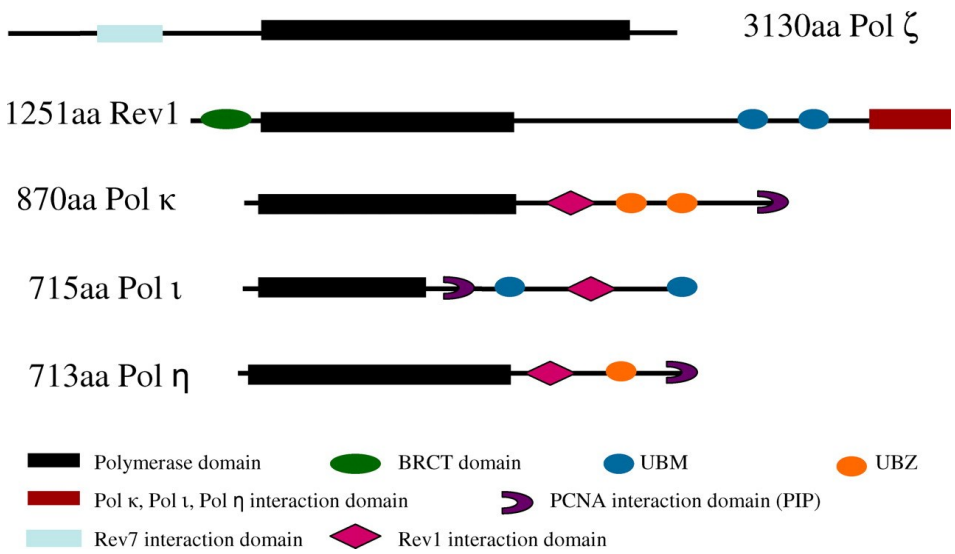

Figure 17: Cartoon representation of the protein domains in human TLS polymerases.

aa: amino acids. Adapted from Waters LS et al. (154)

The N-termini of the Y-family polymerases consist of a conserved polymerase domain of 350-450 residues, which is distinct from common replicative polymerases (replicases) (160, 161). A hallmark of TLS polymerases is a more spacious active site compared to replicases (162). The smaller thumb and finger domains of TLS polymerases give rise to a solvent-exposed active site (159), which is not nearly as constrained as those in the replicases. Besides making fewer contacts with the base pair that needs to be formed (163), the Y-family polymerases lack the O helix of replicases that rotates upon binding of a dNTP to sterically inspect the base pair formation (154). Also, Y-family polymerases lack the intrinsic 3′-5′ exonuclease domain that proofreads the newly replicated strand in many high-fidelity replicases (164). In addition, certain TLS polymerases, such as Rev1 and Pol ι do not use canonical Watson-Crick base

pairing (160, 165). Due to these special features of TLS polymerases, they exhibit error rates in the range of one in every 10,000 bases (as compared to one in every 10^6-10^8 bases for replicases) (164, 166). Therefore, TLS polymerases must be under tight regulation in order to achieve a balance between genetic instability and mutation rates.

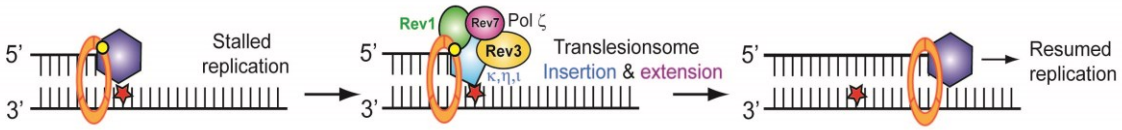

Figure 18: "Polymerase switch" during translesion synthesis

When DNA damage stalls DNA replication, PCNA (orange) is monoubiquitinated (yellow for ubiquitin moiety), which recruits TLS polymerase Rev1 (green) to replace replicase (purple). Rev1 further serves as loading dock to recruit "inserter" (blue) and "extender" (pink and yellow, Rev7 and Rev3) together to bypass the DNA damage (167).

The current widely accepted model for TLS is called "polymerase switch" (168-170) (Figure 18). In this model, when replication is stalled at the DNA damage site, the proliferating cell nuclear antigen (PCNA) (171-173) is monoubiquitinated, which recruits one or more TLS polymerases in a translesionsome complex (167). Many of the TLS polymerases have a basal affinity for unmodified PCNA in vitro through their PIP domain (174-178) or the BRCT domain, in the case of Rev1 (179) (Figure 17). Thus, it is remains to be determined if the monoubiquitination of PCNA serves primarily to signal the need for TLS, serves to strengthen the interaction between PCNA and TLS polymerases via their ubiquitin-binding domains (Figure 17), and/or facilitates the switch between the replicases and the TLS polymerases in a translesionsome complex

(169). Once the inserter polymerase incorporates one or more nucleotides at the DNA damage site, the extender polymerase elongates the DNA chain to bypass the damage site. Afterwards, the replicase is switched back in to replace TLS polymerases for normal replication. Insertion determines the accuracy and mutagenic specificity of the TLS reaction, and is carried out by one of Y-family TLS polymerases such as Pol η, Pol κ, Pol ι or Rev1. In contrast, extension is carried out primarily by Pol ζ (180).

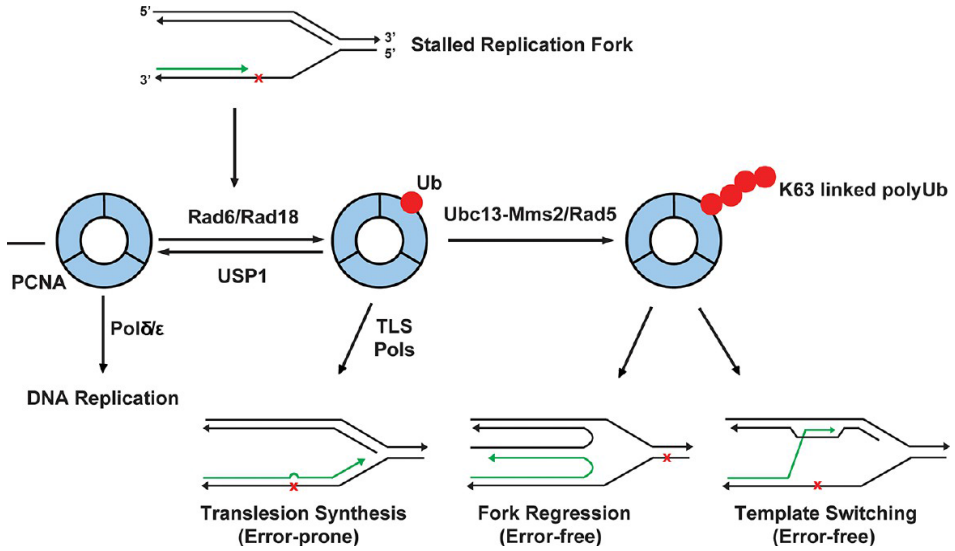

Figure 19: PCNA ubiquitination and TLS

When DNA damage occurs, PCNA is monoubiquitinated at K164 by Rad6 and Rad18, which promotes the recruitment of TLS polymerases to the DNA damage site for TLS. Monoubiquitinated PCNA can be further polyubiquitinated by Ubc13-Mms2/Rad5 with a K63-linked polyubiquitin chain which triggers error-free lesion bypass, including fork regression and template switching. Ubiquitination of PCNA is a reversible process, in which ubiquitin on PCNA can be removed by USP1 in human cells. Adapted from Yang et al (181)

3.1.1.2 Ubiquitination in TLS

PCNA monoubiquitination is the key step of TLS activation. Following DNA damage that stalls the progression of the DNA replication fork, the highly conserved K164 of PCNA is monoubiquitinated by the E2 Rad6 and the E3 Rad18 (171, 182) (Figure 19). PCNA monoubiquitination has been reported in a wide variety of organisms including yeast, Xenopus, chicken, and mammals. The uncoupling of the stalled replicative polymerase and the MCM helicase results in production of single-stranded DNA exposed in the vicinity of the stalled fork. This single-stranded DNA is coated by Replication Protein A (RPA), which in turn interacts with Rad18 and directs the Rad6/Rad18 complex to the site of DNA damage, triggering PCNA monoubiquitination (183, 184).

Besides monoubiquitination, PCNA can also be further polyubiquitinated in the K63-linkage type at K164 by the E2/E2 variant Ubc13/Mms2 and the E3 Rad5 in many organisms including yeast (185). Rad18 recruits Rad5, which further recruits Ubc13/Mms2 through the RING finger domain of Rad5 (186-188). The addition of this chain activates the error-free template switching mechanism of DNA damage bypass (Figure 19) by dislodging TLS polymerases from PCNA, preventing the interaction of the TLS polymerases with PCNA, or recruiting factors that carry out template switching processes (189, 190). The mechanism of template switching is still poorly understood and is hypothesized to proceed by a ‘‘chicken foot’’ mechanism (191) or a homologous

recombination-like mechanism whereby the sister duplex is invaded by a single stranded gap, forming a D-loop (171, 189). In mammalian systems, PCNA polyubiquitination occurs at much lower levels (192). RAD18 and UBC13 are required for PCNA polyubiquitination in human cells, but MMS2 appears to be dispensable (193).

Deubiquitination of PCNA acts as a safeguard against error-prone TLS. USP1 is the deubiquitinase (DUB) that removes the monoubiquitin from PCNA (194). Upon DNA damage, the level of USP1 decreases by an autocleavage event followed by proteasomal degradation of the cleaved products.

TLS polymerases can also be ubiquitinated. Mammalian Pol η, Pol ι, Pol κ, and Rev1 have all been found to be ubiquitinated *in vivo*, modifications that depend on the presence of their intact UBDs and are Rad6-and Rad18-independent (51, 181, 195). The ubiquitination of these polymerases does not require an E3 enzyme, suggesting that the UBDs within the polymerases function as their own auto-ubiquitinating E3 ligases (195, 196). This is not unusual, as the ubiquitination of UBD-containing proteins, in a UBD-dependent manner, has been previously observed (10). Pol η is monoubiquitinated at four lysine residues close to its C-terminus (K682, K686, K694, and K709) (197), while the ubiquitinated residues in Pol ι remain to be determined. Monoubiquitination of Pol η precludes its interaction with PCNA, and upon DNA damage, Pol η is deubiquitinated. These observations support a regulatory role of Pol η monoubiquitination in DNA damage response. Rev1 can also be monoubiquitinated. A recent study showed that

monoubiquitinated Rev1 can interact with FAAP20 of the Fanconi anemia core complex, which recruits the TLS machinery to DNA interstrand crosslink (ICL) sites (198).

3.1.1.3 Ubiquitin-binding domains (UBDs) in TLS

Monoubiquitinated PCNA is recognized by TLS polymerases such as Pol η, Pol ι, Pol κ, and Rev1 via their ubiquitin-binding domains (UBDs), such as the ubiquitin-binding motif (UBM) and ubiquitin-binding zinc finger (UBZ) in conjunction with motifs such as the PCNA-interacting protein (PIP) boxes that recognizes PCNA. Both the UBM and UBZ classes have been discovered through bioinformatics analysis of several yeast two-hybrid screens aimed at finding unconventional ubiquitin interactors (51). In TLS polymerases, the UBDs play an essential role in mediating the interaction between TLS polymerases and the monoubiquitinated PCNA and enhance their binding affinity. (172, 195, 199-204).

Pol ι and Rev1 contain two conserved UBM domains located in their C-terminal regions (14, 195). In Pol ι, both UBM domains are functional in binding ubiquitin and are important for localization of Pol ι to sites of DNA damage and replication foci (51). In mouse REV1, both UBM1 and UBM2 are found to be required for ubiquitin binding, and deletion of either domain greatly decreases the formation of nuclear foci of REV1 in cells (195). Interestingly, the yeast REV1 requires only UBM2 for ubiquitin binding and for its cellular function (64, 195, 205, 206).

Pol η contains one UBZ domain, while Pol κ contains two tandem UBZ domains in the C-terminal region. Mutations in the Pol η UBZ domain that disrupt the UBZ interaction with ubiquitin *in vitro* also greatly reduced the nuclear foci formation of Pol η in response to DNA damage *in vivo* (51, 200). Similar observations were made for Pol κ, again suggesting the important role of PCNA monoubiquitination in recruiting the TLS polymerases to DNA damage sites. The UBZ domain in the human Pol η was shown to be a classical C2H2 zinc-finger domain with a zinc ion sandwiched between an α-helix and two antiparallel β-strands through coordination with the conserved cysteine and histidine residues. The binding interface was mapped onto the α-helix of UBZ and the I44 hydrophobic patch on ubiquitin based on the NMR chemical shift perturbation study (68). Intriguingly, the UBZ domain in *S. cerevisiae* Pol η shows a divergence from the conserved CXXC motif with the two cysteine residues juxtaposed next to each other in the primary sequence. Studies have shown that the yeast Pol η UBZ can bind ubiquitin in a zinc-independent manner (207, 208). The structure of the tandom UBZ domain of Pol κ and its ubiquitin binding specificity remains to be elucidated.

3.1.2 The UBM1 domain of Human Pol ι

Human Pol ι has very low processivity (209). It is able to insert bases opposite the lesion site, but is not able to extend synthesis beyond the inserted base. Human Pol ι is also a low-fidelity enzyme that incorporates dGMP opposite T 3 to 10 times more frequently than the correct nucleotide (210, 211). It also has a 5'- deoxyribose lysase

activity, suggesting a role in base excision repair (BER) in protecting cells from oxidative damage (212). In both BER and TLS, the recruitment of Pol ι requires both its nuclear localization site (NLS) and UBMs (200, 212). It is, like Pol η, localized in replication factories and physically interacts with Pol η (200). Its recruitment to the DNA damage site depends on Pol η (213). Interestingly, Pol ι is even more mobile than Pol η with a shorter $t_{1/2}$ (203). No human condition has been found to be associated with Pol ι deficiency. However, the 129 strain of mice, widely used in the generation of knockout mice, has an ochre mutation close to the N-terminus of the protein and appears to make nonfunctional protein (214). To date these mice have no reported unusual phenotype.

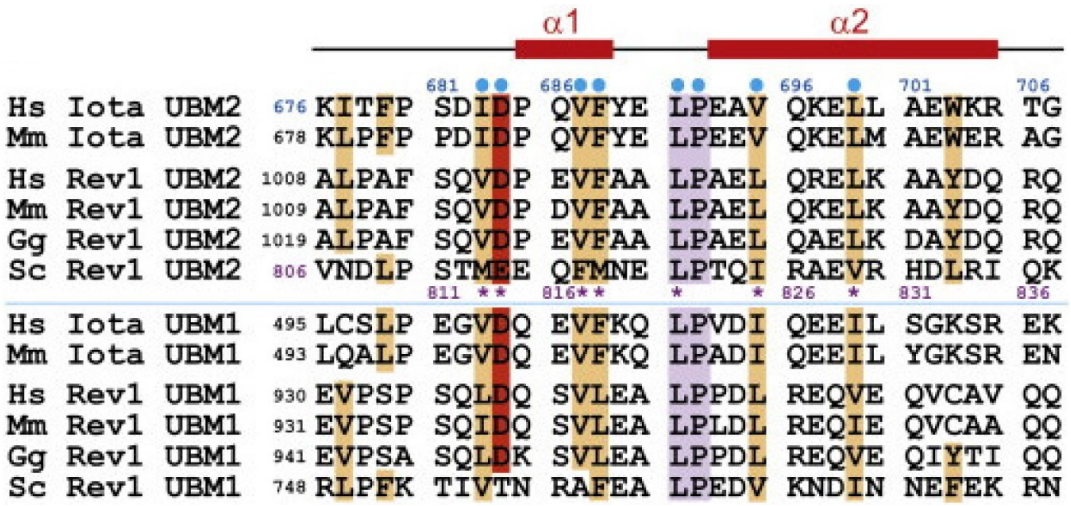

Figure 20: Sequence alignment of UBM domains

Conserved residues are highlighted, with the signature Leu-Pro motif in purple, hydrophobic residues in brown, and negatively charged residues in red. Adapted from Bomar et. al. (64)

The UBM domains were originally identified by their unique ability to interact with an I44A ubiquitin mutant in a yeast two-hybrid screen and were predicted to be

composed of two helical segments with an invariant "Leu-Pro" motif in between (51) (Figure 20). Pol ι contains two copies of UBM within the C-terminal region. Both of the UBMs in Pol ι bind to ubiquitin and are necessary for the accumulation of Pol ι to damage-induced replication foci (51, 200). Mutations within the UBMs of the polymerases do not alter the catalytic activity of the polymerase or its interaction with unmodified PCNA or other polymerases, but they impair their ability to function *in vivo*, possibly due to an affected affinity for binding to monoubiquitinated PCNA or other ubiquitinated target (51, 200, 206).

The structure of human Pol ι UBM2 in complex with ubiquitin has been determined, but the molecular details of the related UBM1 domain remain to be elucidated. To further understand the molecular mechanism of Pol ι UBM-ubiquitin interaction and compare it with UBM2-ubiquitin structure, we applied 4-D fast NMR technique to determine the complex structure of the human Pol ι UBM1-ubiquitin complex.

3.2 Results

3.2.1 Characterization of Pol ι UBM1 and its complex with ubiquitin

3.2.1.1 Interaction identification and binding affinity measurement of hUBM1-ubiquitin complex

The chemical shifts of protein residues are sensitive to subtle differences in the environment, including temperature, pH, and binding partners. Upon binding of the two proteins, the environment of the protein interface changes, affecting the chemical

shifts of the nuclei involved in this association. Therefore, NMR titration can be used to map protein interfaces by titrating an unlabeled interaction partner into a labeled sample and recording the resulting chemical shift perturbations. The addition of unlabeled ubiquitin resulted in the chemical shift perturbation for a subset of resonances of the UBM1 domain, not the GB1 domain, indicating that the human Pol ι UBM1 domain is a functional ubiquitin-binding domain (Figure 21). The shifted peaks suggested that the UBM1-ubiquitin interaction is within the fast exchange regime, characteristic of most moderate-to-weak-affinity interactions.

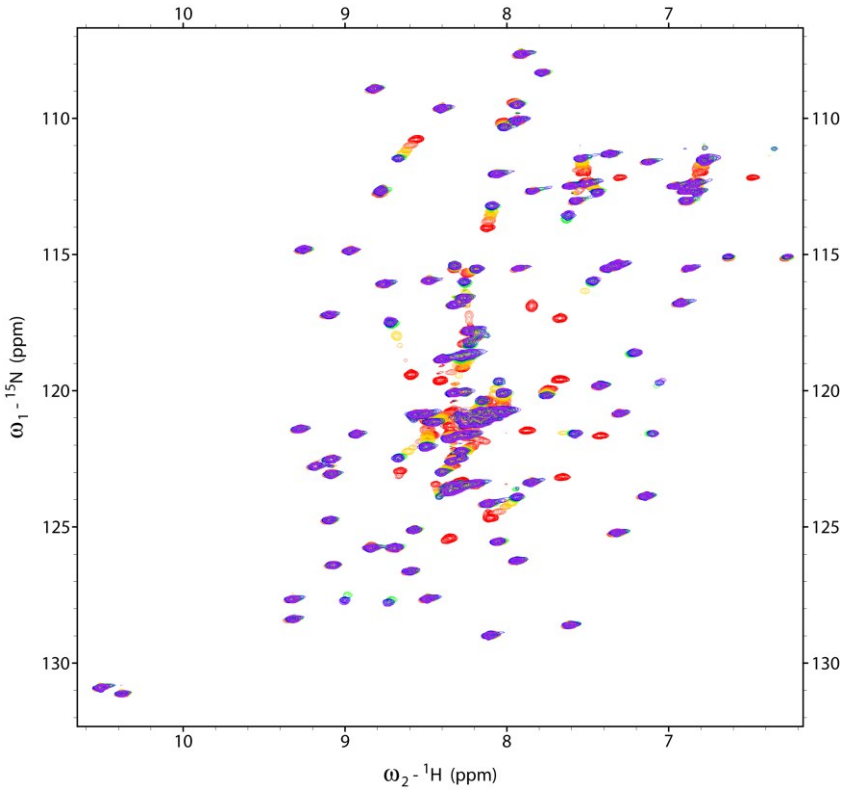

Figure 21: NMR titration of human Pol ι UBM1 domains with ubiquitin

^{1}H-^{15}N HSQC spectra of human pol ι UBM1 with increasing amounts of ubiquitin. Initial (apo state) and final (complex state) points of titrations are depicted in red and purple.

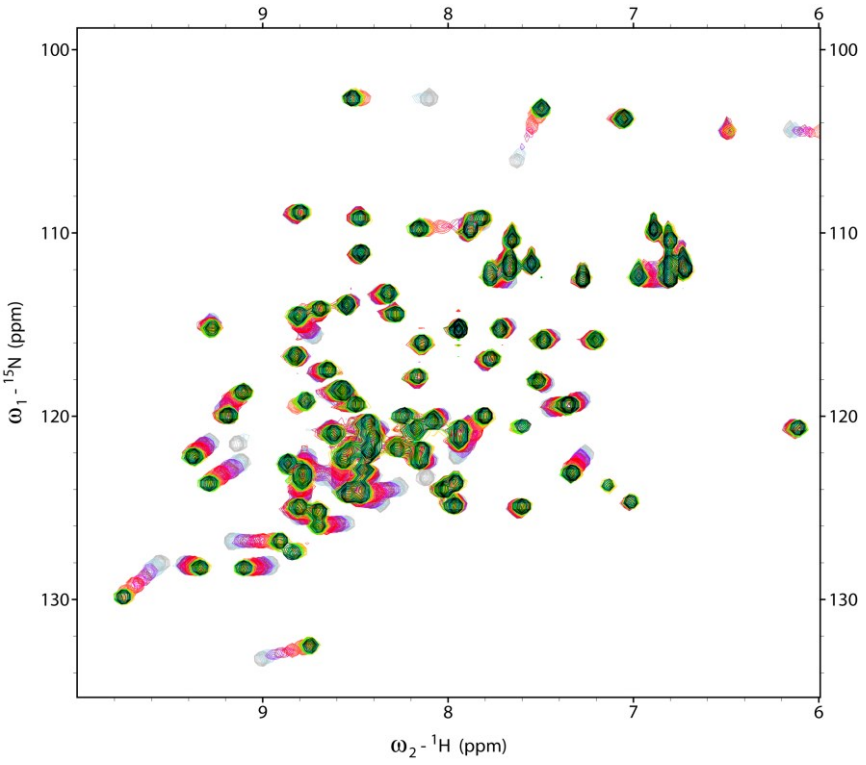

Figure 22: NMR titration of ubiquitin with human Pol ι UBM1 domains

^{1}H-^{15}N HSQC spectra of ubiquitin with increasing amounts of human pol ι UBM1. Initial (apo state) and final (complex state) points of titrations are depicted in grey and black.

Furthermore, the binding affinities of a protein-protein interaction can be quantified based on the chemical shift perturbations. With the addition of unlabeled GB1-fused UBM1 into the ^{15}N-labeled ubiquitin sample (Figure 22), the chemical shifts of a subset of residues of ubiquitin were perturbed. The first and last points of the titration are considered to be the unbound and bound states, respectively. In Figure 23A, the chemical shift corresponding to the free state of a representative amino acid I13 of the ubiquitin is shown in grey, and the saturated or bound state is depicted in black. These two points were connected by a series of titration steps; the normalized chemical shift

changes were calculated for each molar ratio and then used to determine the K_d value of the protein-protein interaction. When unlabeled GB1-fused UBM1 was added to an ^{15}N-labeled ubiquitin sample, resonances from residues V5, K6, T7, T9, K11, I13, I44, A46, L50, L67, V70 of the ubiquitin were significantly perturbed and the perturbations were used for binding-affinity calculation, resulting a K_d of 130±36 μM (Figure 23B). The majority of these perturbed residues are located on the highly conserved, hydrophobic, concave surface of ubiquitin defined by residues L8, I44, and V70 (Figure 1).

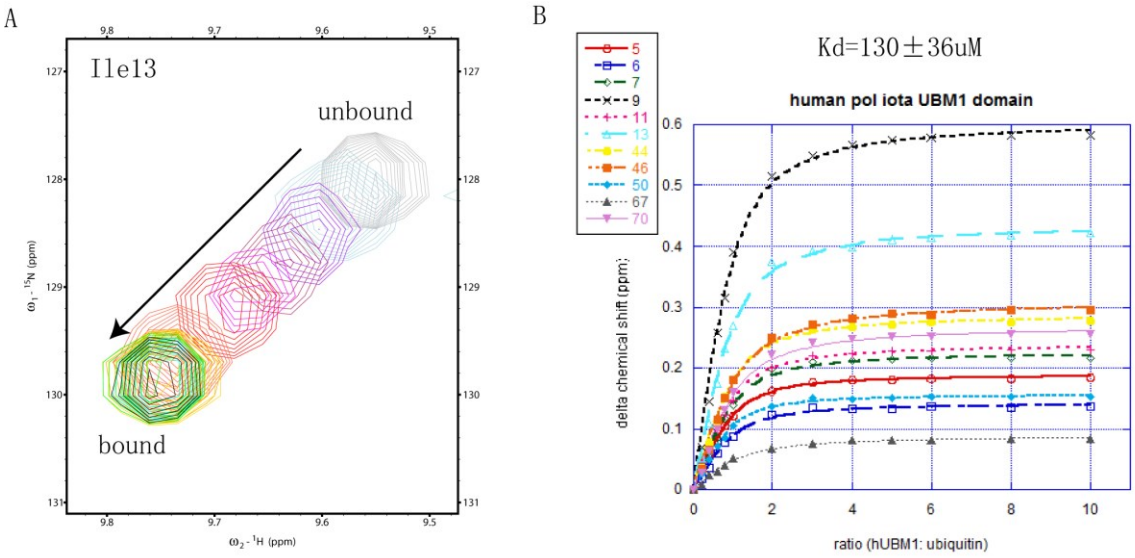

Figure 23: Chemical shift perturbation of ubiquitin titrated with human Pol ι UBM1

Determination of the binding affinity between the human Pol ι UBM1 domain and ubiquitin by NMR titration. (A) A section of the ^{1}H-^{15}N HSQC spectrum depicting changes of the amide resonance for a representative residue, Ile13 of ubiquitin as the molar ratio of UBM1 to ubiquitin increased (grey to black). (B) Determination of the dissociation constant (Kd) by NMR. The chemical shift perturbations for the backbone amides of 11 residues (V5, K6, T7, T9, K11, I13, I44, A46, L50, L67 and V70) of ubiquitin were used to determine the binding affinity of the human Pol ι UBM1 domain toward ubiquitin.

3.2.1.2 Backbone dynamics by heteronuclear NOE experiments

In addition to structure determination, NMR can be used to study protein dynamics on a wide time scale, including characterizing the backbone dynamics of proteins in solution. For example, the local dynamics of a protein can be characterized by the $^{1}H–^{15}N$ heteronuclear NOE experiment (215), which is sensitive to motions on a picosecond to nanosecond time scale. A pair of experiments is recorded with or without the H spin being saturated. The intensities of the resonances in the two experiments can be compared to provide information about the motion of individual N-H bond vectors. The higher the saturated/unsaturated ratio is, the more rigid the backbone. On the other hand, lower or negative values suggest increased local flexibility of the backbone.

A ^{1}H-^{15}N heteronuclear NOE experiment was collected for free GB1-fused UBM1 and GB1-fused UBM1-ubiquitin complex. Except for regions at their termini and the linker region between the GB1 tag and the UBM1 domain, both ubiquitin and UBM1 are quite rigid (Figures 24).

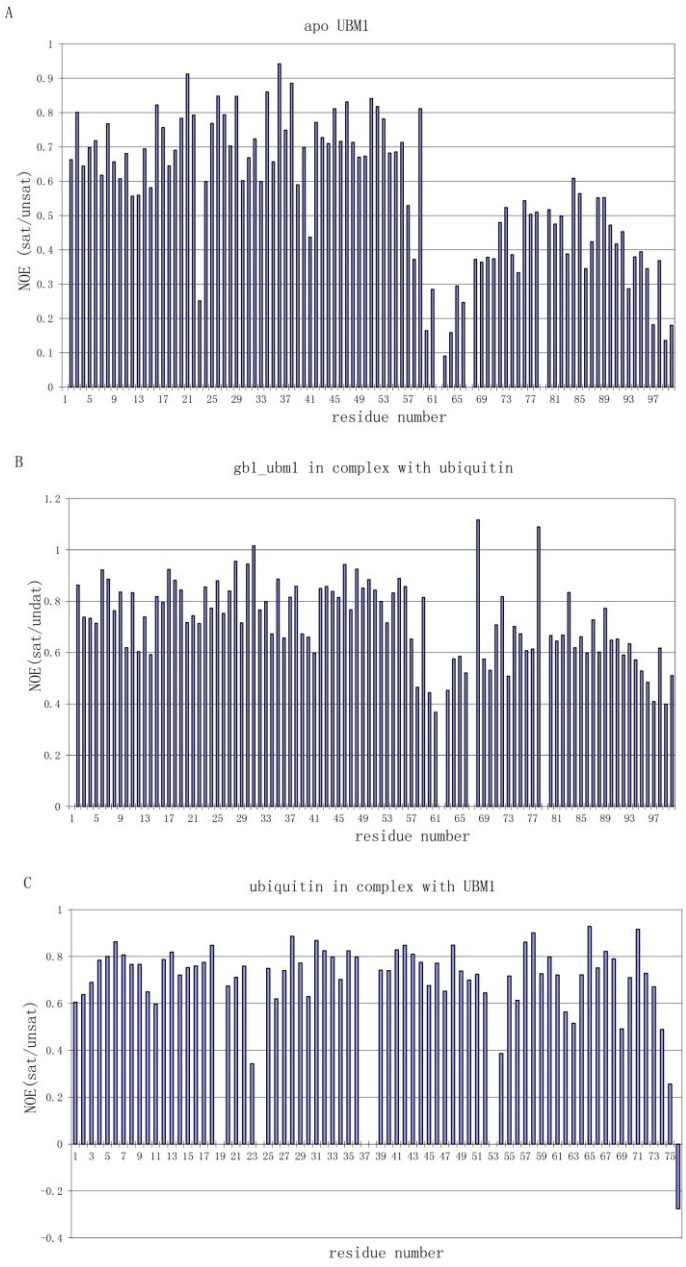

Figure 24: ^{1}H-^{15}N steady-state NOE data for human pol ι UBM1 and ubiquitin

^{1}H-^{15}N Steady-State NOE data for GB1-UBM1 free (A) and bound to ubiquitin (B) and for ubiquitin bound to UBM1 (C). Y-axes show the intensity ratio of the saturated to unsaturated experiment. X-axes indicate the residue number of GB1-UBM1 or ubiquitin, respectively. The linker region between GB1 and UBM1 is composed of residues 57 and 58.

The steady-state NOE data, together with the NMR titration of the GB1-fused UBM1 protein with ubiquitin, indicated that the GB1 tag and UBM1 domain are separate entities and connected by a flexible linker. The GB1 tag is not bound to ubiquitin, nor does it interfere with the ubiquitin-UBM1 binding interface. The establishment of linker flexibility between the solubility tag and fusion protein and the lack of interference by the solubility tag are two critical prerequisites for further structural studies on fusion-tagged proteins.

3.2.1.3 Determination of the oligomeric state by analytical ultracentrifugation

Analytical ultracentrifugation experiments were performed to determine the oligomeric state of the UBM1 domain at various protein concentrations. Figure 25 shows a nonlinear least squares fit of the curve generated by sedimentation equilibrium. The molecular mass of UBM1 was in the range of 11.5 +/- 0.5 kDa at concentrations of 20, 40 and 80 μM as compared with an anticipated monomer molecular mass of 12.0 kDa. This suggests that the UBM1 domain is mainly monomeric in solution under these conditions, which are identical to those used in structure determination and binding studies.

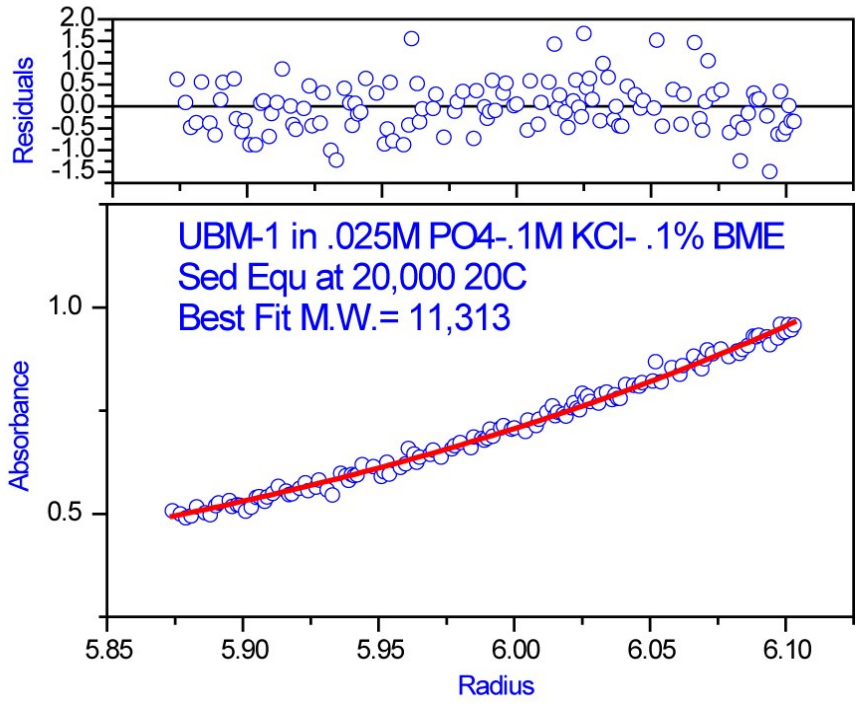

Figure 25: UBM1 is a monomer in solution

A pattern of the analytical ultracentrifugation data is shown for a 80 μM protein solution run at 20,000 rpm and at 20 °C. The best fit molecular weight was 11,313 Da (expected 11,977 Da).

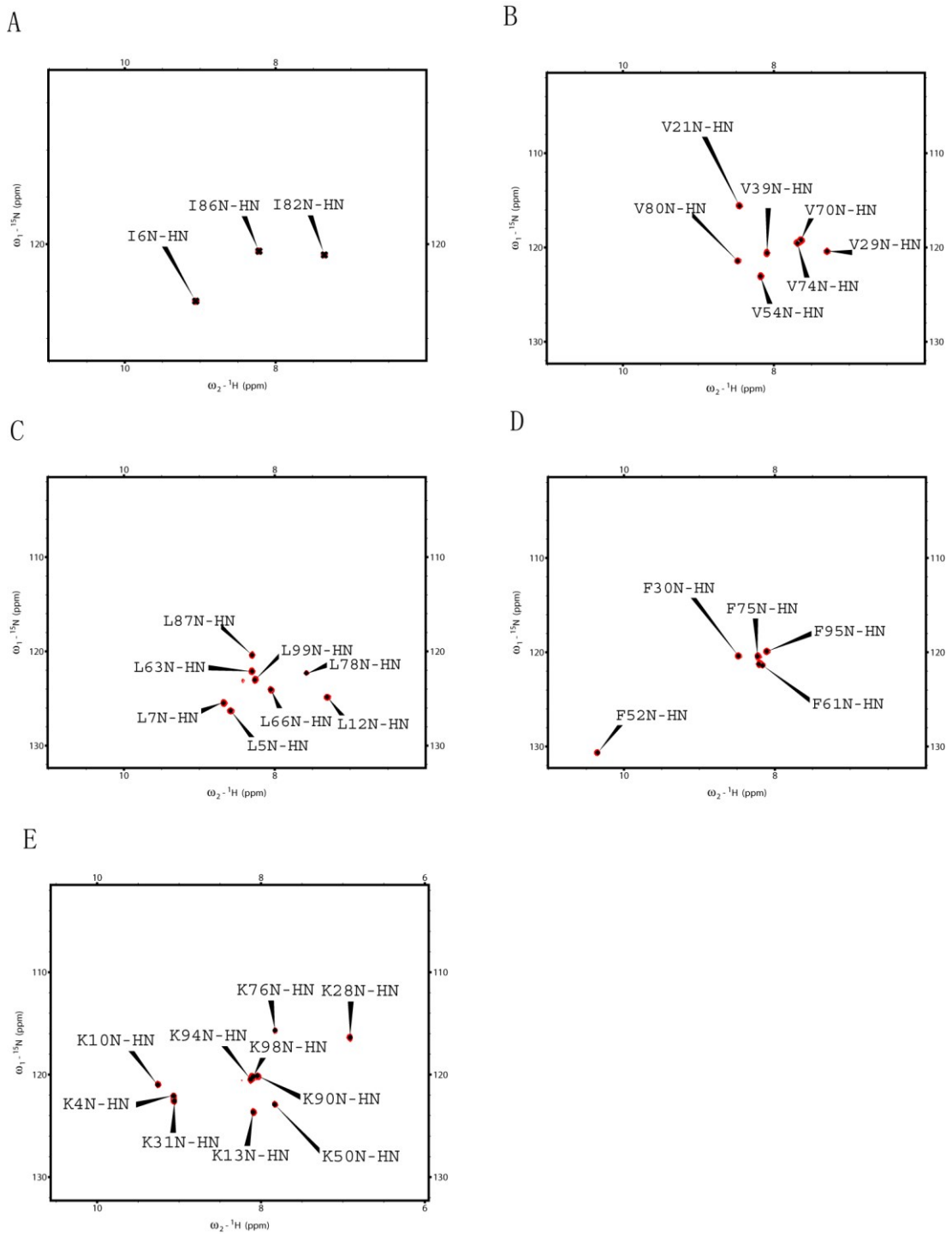

Figure 26: Amino acid specific labeling of GB1-UBM1

^{1}H-^{15}N HSQC spectra for each amino acid specific labelled GB1-UBM1 sample with peaks assigned, including Ile (A), Val (B), Leu (C), Phe (D) and Lys(E).

3.2.2 Solution structure of human Pol ι UBM1 domain

NMR was used to elucidate the solution structure of the human Pol ι UBM1 complex(Figure 27). 4-D sparsely-sampled HNCACB and HN(co)CACB experiments were used for the backbone assignment. However, due to the severe overlapping of the signals in the center area of the ^{1}H-^{15}N-HSQC spectrum, the sequential assignment could not be completed. Therefore, we used the amino acid type specific labeling method to obtain the ^{1}H-^{15}N-HSQC spectrum for certain types of amino acid residues, including Lys, Val, Ile, Leu, and Phe (Figure 26) (216). These individual subsets of ^{1}H-^{15}N resonances provided clues for the backbone assignment while also serving as cross checks to confirm the existing assignments from the 4-D data. 3-D triple resonance experiments were also collected to confirm the backbone assignments and obtain the CO resonance information (217). 4-D sparsely-sampled HC(co)NH and HCCH TOCSY experiments were used for the side chain assignments. 4-D sparsely-sampled ^{15}N- and ^{13}C-separated NOESY-HSQC experiments were used to generate distance constraints. Residual dipolar couplings ($1D_{HN}$, $1D_{HaCa}$) were measured to generate orientation constraints.

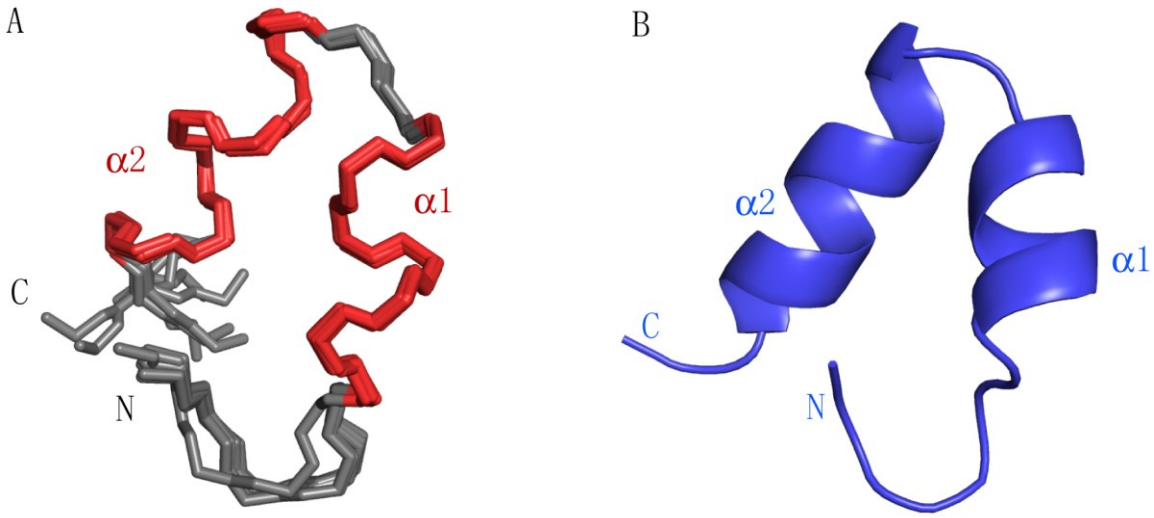

Figure 27: Solution structure of human pol ι UBM1

Structure of human pol ι UBM1. (A) Backbone traces of the NMR ensemble of 10 structures. Helices and loops are colored in red and grey respectively. (B) Cartoon representation.

Except for residues at the termini, human Pol ι UBM1 is well structured (Figure 27), with a mean pair-wise RMSD of 0.35 Å and 1.06 Å for the backbone and heavy atoms of residues 498-520, respectively. The structure of UBM1 consists of two amphipathic helices that are supported by an N-terminal loop folding back to form a β-turn structure. The N-terminal loop adopts a typical β-strand conformation, with the side chains of two residues (L498 and P499) interacting with the hydrophobic surfaces of the first helix following it. Helix 1 contains a single turn (Q504 to F507) and is positioned at a sharp angle with respect to the N-terminal loop. The stability of this short helix is greatly enhanced by the presence of the prototypical N-terminal helix cap residues (218) D503, with its sidechain forming hydrogen bonds with the amides of Q504 and V506.

The C-terminal helix (α2) is longer than helix 1, extending from residue V512 to S520. The two helices lie within a plane at an angle of ~50-60 degree, and they are connected by a short loop containing the signature "Leu-Pro" motif poised for interaction with ubiquitin. The two helices are packed together by the interactions of the core aromatic residue F507 of α1 and other hydrophobic residues, including V502 and L510 flanking helix 1, and I514, I518, L519 of helix 2. Notably, an analysis of the UBM1 structure with the DALI server did not identify any known structure other than ones in the UBM family with a similar fold.

3.2.3 Detection and assignment of intermolecular NOEs from 4-D omit NOE spectroscopy

The intermolecular NOEs of the human Pol ι UBM1-ubiquitin complex were determined using the sparsely-sampled 4-D ^{13}C HMQC-NOESY-HSQC omit NOE experiment. Three samples of gb1UBM1-ubiquitin complexes were prepared, including (1) a uniformly $^{13}C/^{15}N$-labeled complex, (2) a ubiquitin selectively labeled complex, and (3) a gb1UBM1 selectively labeled complex, all at 3 mM concentration. Such a high sample concentration is necessary due to the weak binding affinity between UBM1 and ubiquitin to ensure the vast majority of the protein remains in the complex form. Three identical 4-D ^{13}C HMQC-NOESY-HSQC experiments, each taking 43 hrs, were recorded for each sample, using cosine-weighted random concentric shell sampling (134) with 3190 sampling points digitized on a 64x96x64 grid for the c(t1)-h(t2)-C(t3) indirect dimensions, corresponding to 0.8% of sampling points required for Nyquist sampling.

The omit spectrum was generated by subtracting scaled FIDs of component-labeled complexes from those of the uniformly-labeled complex and reconstructed by SCRUB (141). To account for the minor difference of NOE crosspeak intensities in the three samples, it is preferable to slightly over-subtract the component-specific NOE signals so that all of the diagonal signals are negative. This ensures that the remaining positive signals represent genuine intermolecular NOEs, whereas intramolecular crosspeaks are either completely removed or appear as weak negative signals (Figure 28).

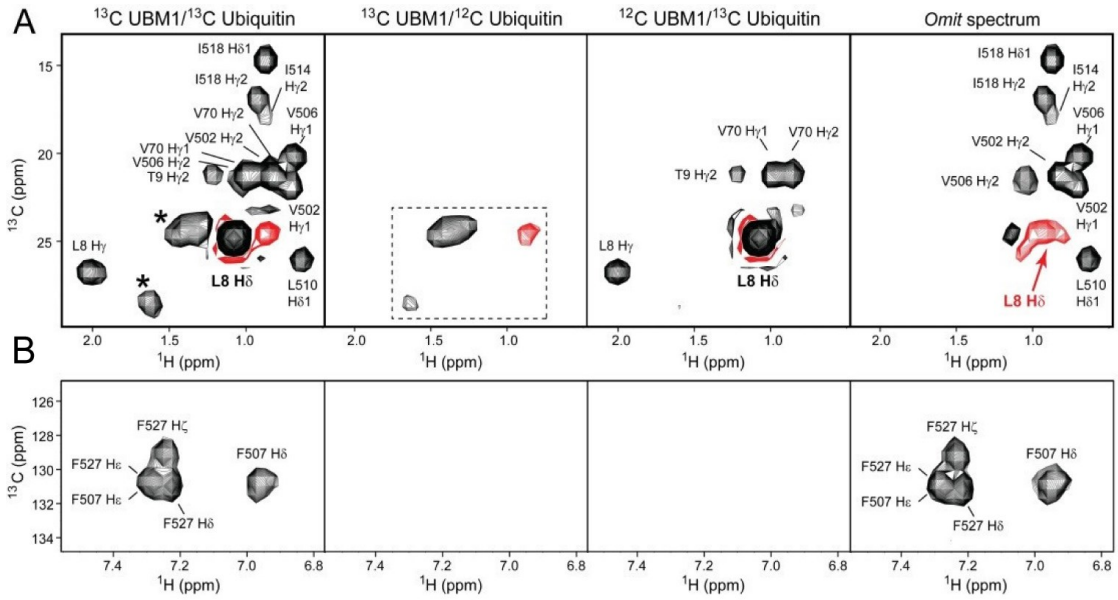

Figure 28: Omit spectrum of the human Pol ι UBM1-ubiquitin complex showing intermolecular NOEs.

4-D ^{13}C HMQC-NOESY-HSQC spectra are collected for the gb1UBM1-ubiquitin complex with both components or with individual components ^{13}C-labeled. Reconstruction of the difference time domain signals of the uniformly labeled protein complex sample from component-labeled samples generates an omit spectrum containing only intermolecular NOEs. Slight over-subtraction of time domain data from individual components generates negative diagonal signals (red) in the omit spectrum and ensures all of the positive cross-peaks originate from intermolecular NOEs. (A) aliphatic regions and (B) aromatic regions show sections of F1-F2 slices of the corresponding 4-D spectra centered at 24.53 ppm in F3 and 1.05 ppm in F4, displaying NOEs to the ubiquitin L8 methyl groups. Boxed peaks or peaks labeled with asterisks are off-plane signals.

A large number of intermolecular NOEs can be observed in the 4-D omit spectrum, providing ample constraints to define the UBM1-ubiquitin binding interface. In particular, numerous intermolecular NOEs between UBM1 residues (V502 from the N-terminal tail, V506 and F507 from the α1, L510 from the turn between the two helices,

I514 and I518 from α2 and F527 from the C-terminal tail) and the L8 methyl groups of ubiquitin are detected in the 4-D omit spectrum (Figure 28).

Of note, the employment of 4-D omit NOE strategy is particularly important in obtaining the complete set of intermolecular NOEs, as many intermolecular NOEs involving L8 methyl groups would be lost in 3-D filtered NOE experiments due the degenerate ^{1}H chemical shifts of the interacting functional groups. The excellent signal separation of the aromatic rings of F507 and F527 is also critical to establish the UBM1 C-terminal tail interaction with ubiquitin (Figure 28B), a distinct feature that separates the UBM1-ubiquitin interaction from the UBM2-ubiquitin interaction.

3.2.4 Solution structure of human Pol ι UBM1-ubiquitin complex

The solution structure of the Pol ι UBM1-ubiquitin complex was determined with 4184 NOEs (including 293 intermolecular NOEs), 190 dihedral angles, and 86 hydrogen-bond constraints. Except for residues at both termini, the UBM1-ubiquitin complex is well structured, with mean RMSD values of 0.34 Å and 0.87 Å for the backbone and heavy atoms, respectively, of residues 496-527 of Pol ι and 1-74 of ubiquitin. A final structural ensemble of 10 structures (Figure 29) was generated with no NOE violations > 0.5 Å and no dihedral angle violations > 5°. The quality and statistics of these structures can be evaluated in Table 2.

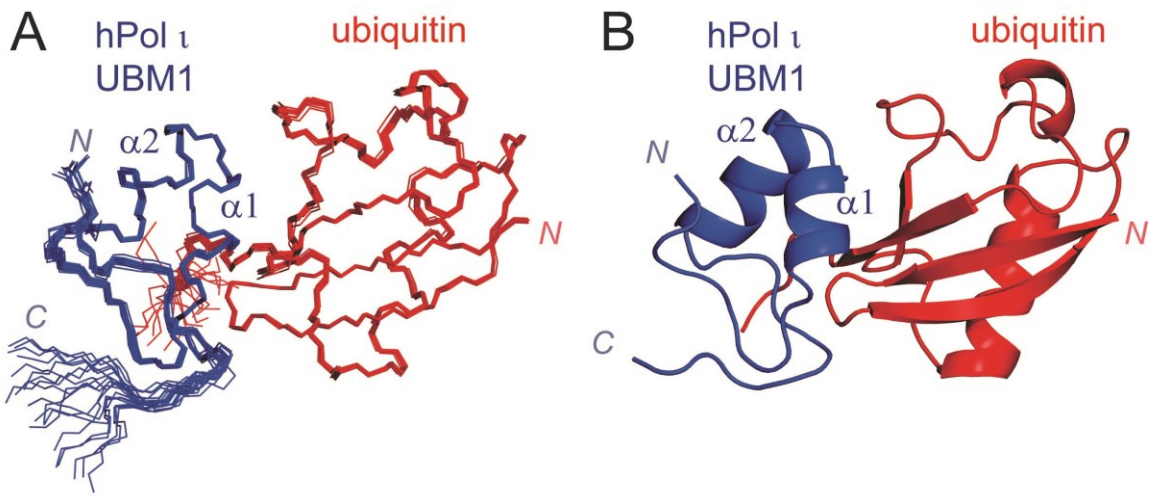

Figure 29: Structure of the human Pol ι UBM1-ubiquitin complex.

(A) Backbone trace of the NMR ensemble of 20 structures. Human Pol ι UBM1 is colored in blue, ubiquitin is colored in red. Two helices of human Pol ι UBM1 are labeled with α1 and α2. (B) Ribbon diagram of the complex, with human Pol ι UBM1 in blue and ubiquitin in red.

Table 2: Structural statistics for human Pol ι UBM1-ubiquitin complex

Structural statistics for the human Pol ι UBM1-ubiquitin complex (20 structures)[a]	
Pol ι UBM1 (62-99)	
NOE distance restraints	898
Intra-residue	210
Sequential	263
Medium-range (1< |i-j| ≤4)	317
Long-range (|i-j| ≥ 5)	108
Hydrogen bond constraints[b]	18
Dihedral angle constraints[c]	66
Ubiquitin (1-76)	
NOE distance restraints	2993
Intra-residue	414
Sequential	628
Medium-range (1< |i-j| ≤4)	634
Long-range (|i-j| ≥ 5)	1317
Hydrogen bond constraints[b]	68
Dihedral angle constraints[c]	124
Intermolecular NOE distance constraints	243
Target function value	0.86 ± 0.12
Ramachandran plot[d]	
Favored region (98%)	94.3
Allowed region (>99.8%)	100.0
Mean pairwise RMSD	
(Pol ι UBM1 64-96; Ubiquitin 1-74)	
Backbone	0.34 ± 0.06 Å
Heavy Atoms	0.87 ± 0.12 Å

[a] None of these structures exhibit distance violations greater than 0.5 Å or dihedral angle violations greater than 5°.

[b] Two constraints ($d_{HN-O} \leq 2.5$ Å and $d_{N-O} \leq 3.5$ Å) are used for each identified hydrogen bond.

[c] Dihedral angle constraints were generated by talos+ based on backbone atom chemical shifts, and by analysis of NOE patterns (108).

[d] MolProbity was used to assess the quality of the structures (219).

The ubiquitin moiety adopts an α/β roll topology with a long α-helix and a smaller 3_{10}-helix packing against the main β-sheet containing five strands. The canonical hydrophobic surface, located centrally on the solvent-exposed surface of the β-sheet and known for interacting with most ubiquitin-binding domains, is recognized by residues from the two helices of UBM1; however, the overall binding area is noticeably shifted toward L8 instead of centering at I44 (Figure 30), a critical residue for most ubiquitin-binding domain (UBD)-ubiquitin interactions (220).

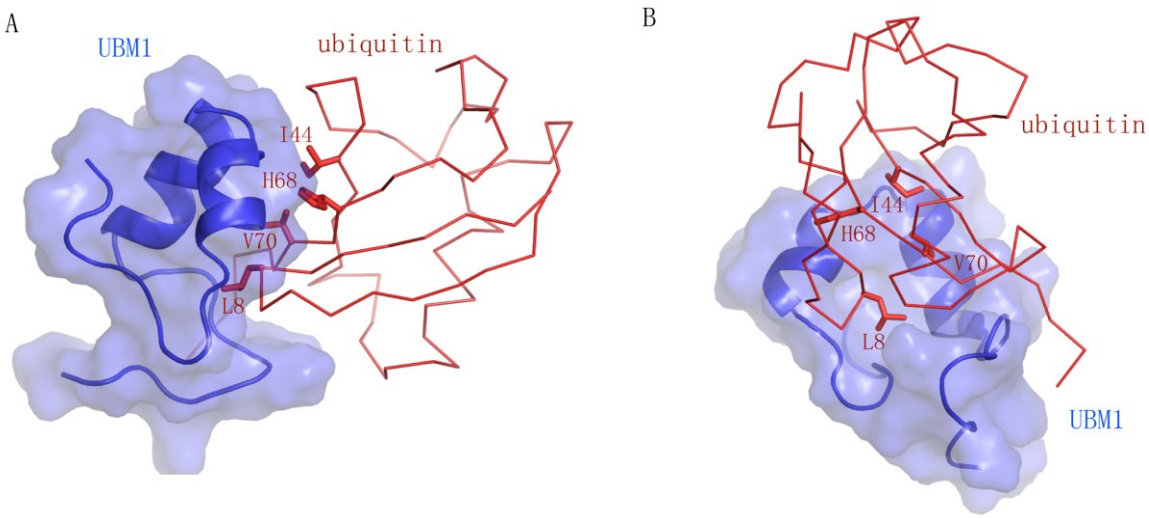

Figure 30: Ubiquitin-binding interface with human Pol ι UBM1

Side view (A) and front view (B) of the ubiquitin-binding interface on UBM1 depicting an interaction network centered at L8, instead of I44, of ubiquitin. UBM1 is depicted in blue and ubiquitin is shown in red. Side chains of UBM1-interacting residues are shown in the stick model on ubiquitin.

A set of hydrophobic residues of UBM1, including V502 located to the N-terminal of helix 1, V506, F507 of helix 1, L510 of the invariant "Leu-Pro" motif connecting helices α1 and α2, I514 and I518 of α2 and F527 of the C-terminal loop form a

network of hydrophobic interactions encircling L8 of ubiquitin (Figure 31). This is supported by numerous 4-D intermolecular NOEs detected between UBM1 and L8 of ubiquitin (Figure 28). This core binding interface is further supported by peripheral hydrophobic interactions formed between V506, L510, P511 and I514 of UBM1 and I44 and the aliphatic side chains of K6 and H68 of ubiquitin. Similar to other UBM domains, the invariant "Leu_{510}-Pro_{511}" motif of UBM1 plays an important role in supporting ubiquitin binding, with the side chains of L510 inserting into the hydrophobic pocket of ubiquitin defined by L8, I44, H68 and V70, and P511 nudging into the shallow surface groove formed by I44, G47 and Q49.

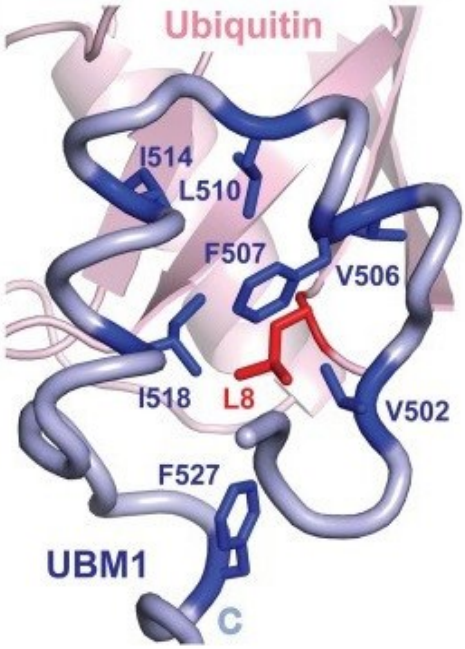

Figure 31: Interface of the human Pol ι UBM1-ubiquitin complex, showing an interaction network centered at L8 of ubiquitin.

UBM1 is depicted in pale blue and ubiquitin is shown in pale pink. Side chains of L8 of ubiquitin is shown in the stick model and colored in red. L8-interacting residues of UBM1 are shown in the stick model and colored in blue.

The interaction between ubiquitin and UBM1 also appears to depend critically on the stability of the N-terminal short helix in UBM1. The N-terminus of α1, including the highly conserved N-cap residue of α1, D503, is located in the vicinity of the ubiquitin surface and is involved in the interaction with ubiquitin. Its amide group forms a hydrogen bond with the carbonyl group of L8 of ubiquitin, thus providing an anchoring point for ubiquitin binding. Such a configuration is well supported by intermolecular NOEs between D503 Hα and Hβ protons of human UBM1 and the Hα3 protons of G10 of ubiquitin (Figure 32)

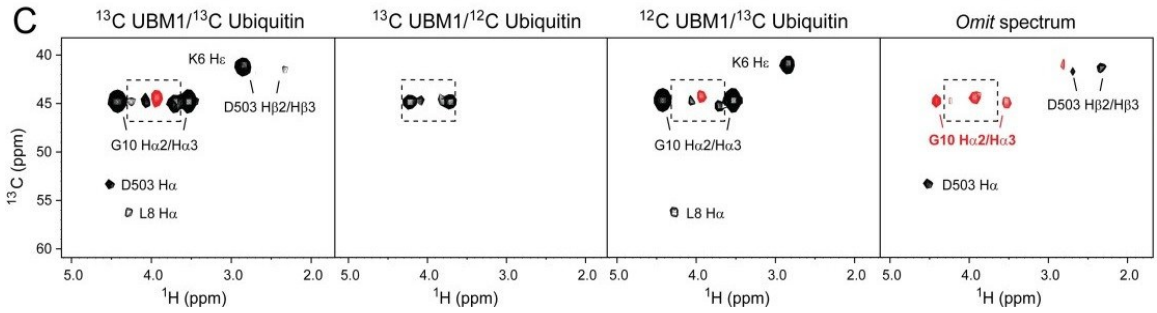

Figure 32: Omit spectrum of the human Pol ι UBM1-ubiquitin complex showing intermolecular NOEs of D503

2-D slices of the 4-D 13C HMQC-NOESY-HSQC spectra centered at 45.07 ppm in F3 and 3.51 ppm in F4, illustrating NOEs to the G10 Hα3 proton of ubiquitin. Boxed peaks are off-plane signals.

3.3 Discussion

3.3.1 Unique ubiquitin recognition by UBM

UBM adopts a unique protein fold that has not been observed in other UBDs previously. In addition, the ubiquitin recognition mode of UBM is also distinct from all the other known UBDs. UBM belongs to the α-helical UBD family, but its interaction with ubiquitin is centered at L8, instead of I44 of ubiquitin as seen for other α-helical UBDs. UIM, MIU/IUIM and DUIM all use a single helix to bind to the solvent-exposed β-sheet of ubiquitin (40-43, 58), with the ubiquitin-binding helix oriented either parallel or antiparallel to the central β-strand of ubiquitin (β5) (Figure 33A). In contrast, UBM binds to ubiquitin using two consecutive helices, with helix 1 being perpendicular to β5 of ubiquitin and helix 2 poised at an angle. The usage of two α-helices has been seen in UBA-ubiquitin recognition (61, 62, 221). However, UBA uses two discontinuous helices,

α1 and α3 of the three-helix bundle, which arranges in an "up-up" topology, whereas the two α-helices of UBM are in an "up-down" topology (Figure 33B). The GAT domain also use α1 and α2 to interact with ubiquitin (47, 222), however, they are almost antiparallel with each other, whereas two α-helices of UBM are at angles of about 50-60° on the same plane (Figure 33C). Therefore, UBM utilizes a distinct ubiquitin recognition mode.

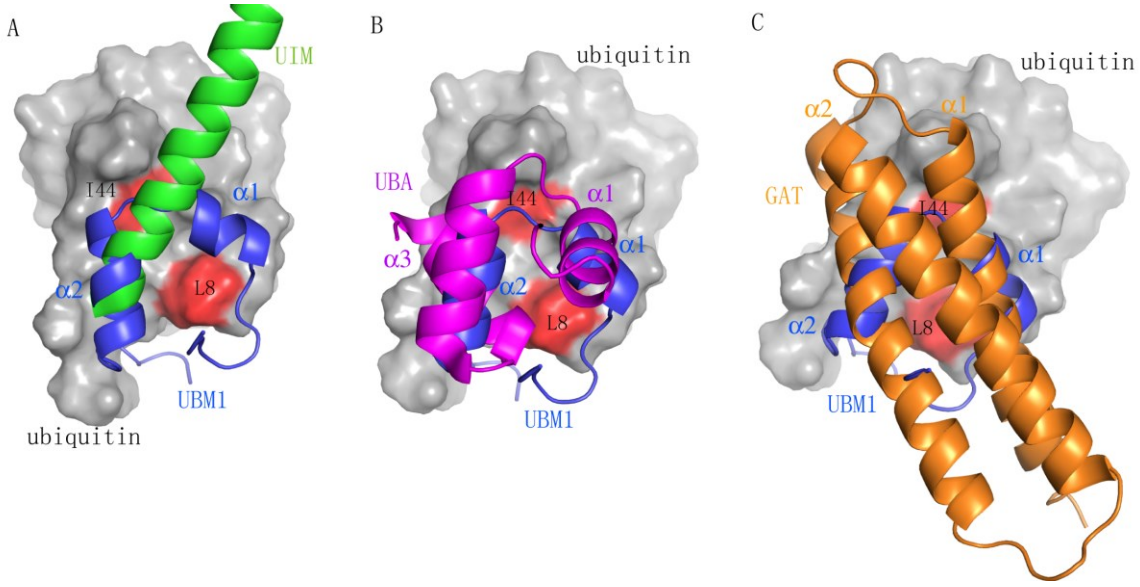

Figure 33: A unique mode of ubiquitin recognition by UBM1

(A) Ubiquitin binding by UIM (green, PDB 1YX5) occurs via a single helix. Ubiquitin binding by UBA (magenta, PDB 2JY6) and GAT (orange, PDB 1YD8) as representative domains containing three-helix bundles is shown in panel B and C, respectively. UBM is colored in blue. Ubiquitin is in grey surface mode.

3.3.2 Structure comparison of UBM homologs

Compared with human Pol ι UBM2-ubiquitin complex structure (64), human Pol ι UBM1-ubiquitin complex adopts an overall similar fold and shares similar interface

with ubiquitin. The two structures overlay fairly well, with r.m.s.d. of 0.8 Å (Figure 34A). They both contain two α-helices with almost identical angles. The "Leu-Pro" motif between the two helices plays an important role in interacting with ubiquitin. The major difference in the UBM portion is that the second helix in UBM1 is one helix turn shorter than in UBM2. This is due to the sequence difference between UBM1 and UBM2. The replacement of conserved Trp in UBM2 with Lys in UBM1 disrupts the second helix, which makes the hydrophobic core of UBM1 smaller than UBM2. The N-terminal loop in both structures adopts a similar turn. In UBM2-ubiquitin complex structure, the C-terminal tail is disordered; however, the C-terminal loop of UBM1 folds back and extends towards ubiquitin for interacting with the C-terminal part of β5 and the loop containing L8 of ubiquitin (Figure 34A).

Almost at the same time as we were finishing our work on the human Pol ι UBM1-ubiquitin complex structure, the mouse Pol ι UBM1-ubiquitin complex structure was also determined by solution NMR (63). Mouse UBM1 also contains two α-helices for interacting with the hydrophobic patch of L8 of ubiquitin. Interestingly, structural comparison of the human Pol ι UBM1-ubiquitin complex determined by 4-D NOE with the mouse Pol ι UBM1-ubiquitin complex (PDB 2KWV) has revealed larger-than-expected structural deviation at the N-terminus of α1 of UBM1 (Figure 34B). In the human Pol ι UBM1-ubiquitin complex, the N-terminus of α1, including the highly conserved N-cap residue of α1, D503, is located in the vicinity of the ubiquitin surface

and is involved in the interaction with ubiquitin. Such a configuration is well supported by intermolecular NOEs between D503 Hα and Hβ protons of human UBM1 and the Hα3 protons of G10 of ubiquitin (Figure 32). In contrast, in the mouse Pol ι UBM1-ubiquitin complex structure, the N-terminus of α1 is tilting substantially away from ubiquitin, in which the shortest distances between ubiquitin G10 Hα protons (i.e. Hα3) and the Hβ protons of the corresponding Asp in UBM1 (D501 in mouse Pol ι UBM1) are over 7.4 Å. The ubiquitin G10 Hα3 and mouse UBM1 D501 Hα protons are located further apart, with a distance of ~10.5 Å, reflecting a lack of intermolecular NOEs between these protons.

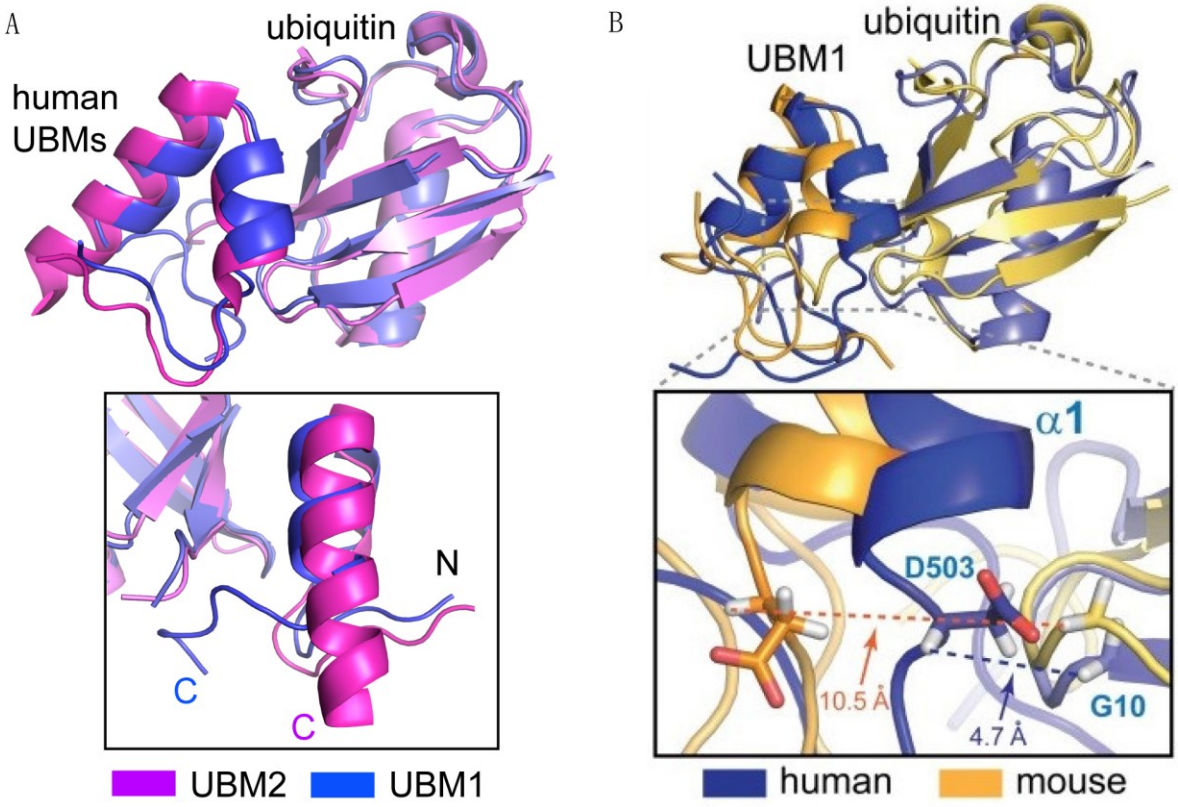

Figure 34: Structure comparison of UBM homologs

(A) overlay of the human Pol ι UBM1-ubiquitin complex (blue, this work) with UBM2-ubiquitin complex (magenta, PDB ID 2KHW). The difference between the C-terminal parts of the two UBM domains is shown in the lower zoomed-in panel. (B) Overlay of the human Pol ι UBM1-ubiquitin complex (blue; this work) with the mouse Pol ι UBM1-ubiquitin complex (yellow; PDB ID 2KWV) by superimposing the ubiquitin part. The zoom-in panel shows the distances between the ubiquitin G10 Hα3 proton and the Hα proton of a conserved UBM1 Asp residue (D503 in human Pol ι UBM1 and D501 in mouse Pol ι UBM1).

3.3.3 Implication on the molecular mechanism of Pol ι UBM-ubiquitin recognition in translesion synthesis

The UBM domains appear to have important effects on Pol ι function. Studies have showed that foci accumulation of Pol ι in S phase of undamaged cells requires two

functional UBMs (64). Deletion of either of the two UBMs of Pol ι abolishes its accumulation at the replication foci. In undamaged cells, the amount of monoubiquitination of PCNA is very limited. Therefore, this data suggests that the UBMs of Pol ι are engaged in recognition of other proteins, most likely in their ubiquitinated form, at replication foci. In this regard, it is worth noting again that besides PCNA, all four mammalian Y-family polymerases can also be ubiquitinated (51, 201, 223) . Therefore, the UBMs may serve as a scaffolding unit to recruit ubiquitinated replication factors, forming a UBM-mediated interaction network for the assembly of the translesionsome, while monoubiquitination of PCNA upon DNA damage can further enhance this assembly process.

Our study shows Pol ι UBM1 and UBM2 adopt a similar ubiquitin recognition mode, which is centered on L8 of ubiquitin instead of canonical I44 of ubiquitin. It will be interesting to elucidate whether this unique ubiquitin recognition mode of the UBM domains is required for normal Pol ι function. When the UBZ domain of Werner helicase interacting protein 1 (Wrnip1) was exchanged with the Pol ι UBM1 domain, foci formation of Wrnip1 was abolished (224). This data suggests that L8-centered ubiquitin interaction by UBM1 cannot be replaced with I44-centered ubiquitin interaction by WRNIP UBZ. Similar domain swapping experiments can be done to illustrate the specific function of Pol ι UBM domains. The Pol η UBZ domain can be swapped in for either UBM1 or UBM2 or both UBM domains of Pol ι to test the importance of L8-

centered ubiquitin recognition in Pol ι function. The two UBM domains of Pol ι can be swapped to test the relative importance of the two UBM domains in Pol ι. This might help explain whether the presence of two UBM domains in Pol ι is simply for enhanced interaction with ubiquitin or whether each UBM domain has a dedicated function.

Another interesting question is whether the specificity of Pol ι UBM domain function depends on the context of ubiquitination. In other words, if the UBM domain can interact with ubiquitinated replication factor and if this interaction is required for Pol ι function, it will add another level of complexity to this highly controlled translesion synthesis process.

3.4 Materials and Methods

3.4.1 Molecular cloning

The DNA sequence corresponding to the UBM1 domain of human DNA Y family polymerase ι (residue 491-530) was synthesized. PCR-amplified DNA was digested and ligated into a modified pET30 vector (EMD Biosciences, Inc.) between the BamHI and XhoI restriction sites. The final construct contains an N-terminal GB1 tag (225) and C-terminal His6 tag. For the construct used in the amino acid type specific labeling, the DNA fragment containing the GB1 tag, the UBM1 sequence and the His tag was digested from the pET30 vector and inserted into pET21a vector. (EMD Biosciences, Inc.) The DNA sequence of human ubiquitin was PCR-amplified from a human cDNA library and cloned into the pET15b vector (EMD Biosciences, Inc.), with an N-terminal His6 tag

and a thrombin cleavage site in between. The presence of the correct inserts of these constructs was confirmed by DNA sequencing.

3.4.2 Protein purification

The GB1-fused human Pol ιUBM1 protein of was overexpressed in *Escherichia coli* BL21(DE3) STAR cells (Invitrogen). For amino- acid type specific labeling, the GB1-fused UBM1 protein was overexpressed in *Escherichia coli* DL39 cells (LeMaster D.M.). Bacterial cells were induced at 20 ºC with 0.25 mM IPTG for 18 hours. Cells from a 1 L culture were collected by centrifugation and lysed by French press in 30 mL of lysis buffer (50 mM phosphate, 300 mM NaCl, 0.1% β-mercaptoethanol, pH=8). Cell lysates were cleared by centrifugation and loaded onto a Ni^{2+}-NTA column (Qiagen). The GB1-fused UBM1 protein was eluted with at least five column volumes of elution buffer (50 mM phosphate, 300 mM NaCl, 250 mM imidazole, 1% β- mercaptoethanol, pH=8). The elution fractions containing protein were concentrated to 2ml with Amicon Ultra Filters (3000 Da cutoff, Millipore). The GB1-fused hUBM1 protein is further purified using size exclusion chromatography (Superdex 75, GE Healthcare) with a buffer containing 25mM phosphate, 100 mM KCl, 1% β–mercaptoethanol, pH=7.

The human ubiquitin was overexpressed in *Escherichia coli* BL21(DE3) STAR cells (Invitrogen). Bacterial cells were induced at 20 ºC with 1 mM IPTG for 18 hours. Cells from a 1 L culture were collected by centrifugation and lysed by French press in 25 mL of lysis buffer (50 mM phosphate, 300 mM NaCl, pH=8). Cell lysates were cleared by

centrifugation and loaded onto a 5ml Ni^{2+}-NTA column (Qiagen). The ubiquitin was eluted with at least five column volumes of elution buffer (50 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH=8). The elution fractions containing protein were concentrated with Amicon Ultra Filters (3000 Da cutoff, Millipore) and buffer exchanged into gel-filtration buffer (25 mM phosphate, 100 mM KCl, pH=7) to remove the imidazole. Five units of thrombin (Haemotologic Technologies, Inc.) were added for digestion to remove the N-His_6 tag. Thrombin was removed by loading the sample onto a 1-mL benzamidine column (GE Healthcare); the protein was collected in the flow through and washed with the gel filtration buffer (100 mM KCl, 20mM phosphate, pH=7). The flow through was added to the Ni^{2+}-NTA column to remove the N-His_6. The untagged ubiquitin was collected in the flow through by rinsing with the gel-filtration buffer. The flow through and washes was concentrated and further purified using size-exclusion chromatography (Superdex 75, GE Healthcare) with the gel-filtration buffer.

3.4.3 Analytical Ultracentrifugation

The molecular weight of the free UBM1 domain in 25 mM phosphate, 100 mM KCl (pH=7.0) was estimated by sedimentation equilibrium at 20,000 rpm at 20 °C on a Beckman model XL-A Analytical Ultracentrifuge. The v_{bar} of the protein and ρ of the solvent at 20 °C were estimated to be 0.727 and 1.0043, respectively, by the SEDNTRP program. The protein was examined at several concentrations (20 μM, 40μM and 80 μM) and monitored by the absorbance at 292 nm. Nonlinear least square curves were

generated by the IDEAL-1 program (Beckman). The analytical centrifugation experiment was performed by Harvey Sage (Duke University).

3.4.4 NMR titration for measurement of binding affinity

The K_d value for the human Pol ι UBM1-ubiquitin complex was calculated based on NMR titration data. Unlabeled GB1 fused-UBM1 was titrated with increasing molar ratios into a 0.5 mM sample of the ^{15}N-labeled ubiquitin. Normalized chemical shift changes (Δδ) were calculated as $\Delta\delta = \sqrt{(\delta_{HN})^2 + (o.2\delta_N)^2}$ for each molar ratio (M).

The dissociation constant (K_d) and value of maximum chemical shift perturbation ($\Delta\delta_{max}$) were extracted from perturbation data of 11 nonoverlapping residues, V5, K6, T7, T9, K11, I13, I44, A46, L50, L67, V70 using the equation:

$$\frac{\Delta\delta}{\Delta\delta_{\max}} = 0.5\left[(M+1+\frac{K_d}{[Ub]_T}) - \sqrt{\left(\left(M+1+\frac{K_d}{[Ub]_T}\right)^2 - 4M\right)}\right]$$

3.4.5 NMR sample preparation and data collection for structure determination

To prepare isotope-labeled sample, bacterial cells were cultured in M9 minimal media using ^{15}N-NH_4Cl and ^{13}C-glucose as the sole nitrogen and carbon sources (Cambridge Isotope Laboratories). For amino acid type specific labeling, bacterial cells were cultured in M9 minimal media using ^{15}N labeled amino acid with unlabeled NH_4Cl and glucose. Purified protein was buffer exchanged into NMR buffer (100mM KCl, 25mM phosphate, pH 7) with either 10%D2O or 100%D2O and concentrate to final

concentration of 3mM for both apo GB1-fused UBM1 and GB1-fused UBM1-ubiquitin complex.

NMR experiments were conducted at 25 °C using Agilent INOVA 600 or 800 MHz spectrometers. 4-D sparely-sample HNCACB and HN(co)CACB experiments and four pairs of sparse-sampled 3-D triple-resonance experiments were collected for backbone resonances assignment. 4-D sparsely-sampled HC(co)NH-TOCSY and HCCH TOCSY experiments were collected for the side-chain resonances assignment. The traditional 3-D ^{15}N-separated NOESY-HSQC experiments using a uniformly ^{15}N-labeled complex sample and the 4-D sparsely-sampled ^{13}C-HMQC-NOESY-^{15}N-HSQC and ^{13}C-HMQC-NOESY-HSQC experiments using a uniformly ^{15}N, ^{13}C-labeled sample were collected for NOE assignments.

NMR data were processed by NMRPIPE (151) and SCRUB (141). NMR spectra were analyzed with XEASY (153) and SPARKY (152).

3.4.6 Generation of 4-D omit spectra and identification of intermolecular NOEs

For intermolecular NOE identification, three identical 4-D ^{13}C HMQC-NOESY-HSQC experiments were recorded for each sample (uniformly ^{15}N, ^{13}C-labeled, ubiquitin selectively labeled and gb1UBM1 selectively labeled). Each experiment used a mixing time of 150ms, with spectral widths of 70ppm for the ^{13}C dimensions (t1 and t3) and 11ppm for the ^{1}H(t2) dimension and t_{max} of 4.54ms for the ^{13}C (t1 and t3) dimensions and 10.85ms for the ^{1}H (t2) dimension. Each experiment was recorded for 43 hr. The omit

spectrum was generated by subtracting scaled FIDs of gb1UBM1-labeled complexes and ubiquitin-labeled complex from those of the uniformly-labeled complex with the scaling factor of 1.07 and 1.05, respectively and reconstructed by SCRUB (141). Intermolecular NOE crosspeaks from the omit spectrum were analyzed manually.

3.4.7 Heteronuclear NOE experiment

The heteronuclear ^{1}H-^{15}N NOE experiments (215) were conducted using a 1mM ^{15}N-labeled GB1-fused UBM1 sample for the measurements of the apo state of UBM1 and using a 2mM GB1-fused UBM1-ubiquitin complex sample for the measurements of UBM1 and ubiquitin in the complex state. The buffer condition was the same as in the NMR structural study. The ^{1}H-^{15}N NOE values were calculated from the ratios of peak intensities in the saturated versus unsaturated spectra.

3.4.8 Residual Dipolar Coupling (RDC) measurement

Residual dipolar couplings ($^{1}D_{HN}$, $^{1}D_{H\langle C\langle}$) of free UBM1 were determined from the difference in couplings between an isotropic and a liquid crystalline Pf1 phage (101) sample (~20mg/ml) or a DMPC/DHPC bicelle sample (7% liquid with q=2.9, 13% cholesterol sulfate) (226). The buffer condition was the same as in the NMR structural study. A 3-D J_{HN}-coupled HNCO TROSY and a 3-D $J_{H\alpha C\alpha}$-coupled HNCO TROSY experiment (227) were used to measure the $^{1}D_{HN}$ and $^{1}D_{H\alpha C\alpha}$ couplings, respectively.

3.4.9 Structure calculation

For the UBM1 apo structure, NOE crosspeaks from 3-D ^{15}N-separated NOESY-HSQC and 4-D sparsely-sampled ^{13}C-HMQC-NOESY-^{15}N-HSQC and ^{13}C-HMQC-NOESY-HSQC were used for automatic CYANA (149) structure calculation and refinement in the presence of orientation constraints derived from the two sets of RDC measurement in two different alignment media and dihedral angle constraints derived from TALOS+ (108) analysis of chemical shift information.

Manually-assigned intermolecular NOE crosspeaks were converted into distance constraints using the calibration module in CYANA (149). NOE crosspeaks from 3-D ^{15}N-separated NOESY-HSQC and 4-D sparsely-sampled ^{13}C-HMQC-NOESY-^{15}N-HSQC and ^{13}C-HMQC-NOESY-HSQC using a uniformly labeled complex sample were used for automatic CYANA (149)structure calculation and refinement in the presence of manually assigned intermolecular NOE constraints and dihedral angle constraints derived from TALOS+ (108) analysis of chemical shift information. The final structural ensembles (20 structures) of the human Pol ι UBM1-ubiquitin complex display no NOE violations > 0.5 Å and no dihedral angle violations > 5 °. The statistics of the structural ensemble is shown in Table 2. The human Pol ι UBM1-ubiquitin complex structure has been deposited in RCSB and BMRB (PDB ID: 2mbb; BMRB ID: 19394).

4. Structural study of human FAAP20 ubiquitin-binding zinc finger domain (UBZ) in complex with ubiquitin

4.1 Background

4.1.1 Fanconi Anemia (FA) pathway

Fanconi Anemia (FA) is a rare autosomal recessive or X-linked genetic disease, which was discovered by the Swiss pediatrician, Dr. Guido Fanconi, in 1927. Clinically, this chromosomal instability disorder is strikingly heterogeneous. Its hallmarks are bone marrow failure, congenital abnormalities, and early onset of cancer. Even though it is a rare disease (the disease affects at minimum 1 in 100,000 individuals (228)), Fanconi Anemia today has become a very well established model system for several biological processes, including DNA repair, cancer progression and protein ubiquitination (229-234).

The hallmarks of FA at the molecular level are the cellular hypersensitivity to a series of DNA damaging agents (235) such as diepoxybutane (DEB), mitomycin C (MMC) or cisplatin, that generate DNA crosslinks, including interstrand crosslinks (ICLs) (236-238). ICLs affect processes that inherently require DNA unwinding and strand separation such as DNA replication and transcription (239). Consequently, ICLs are particularly deleterious if unresolved.

Identification of mutated genes in FA patients has discovered a diverse set of proteins necessary for crosslink repair. Currently, sixteen FANC genes associated with patient mutations (*FANCA-FANCQ*) have been described. They constitute a common

DNA repair pathway, called the Fanconi Anemia Pathway, to resolve the ICLs encountered during replication (240, 241).

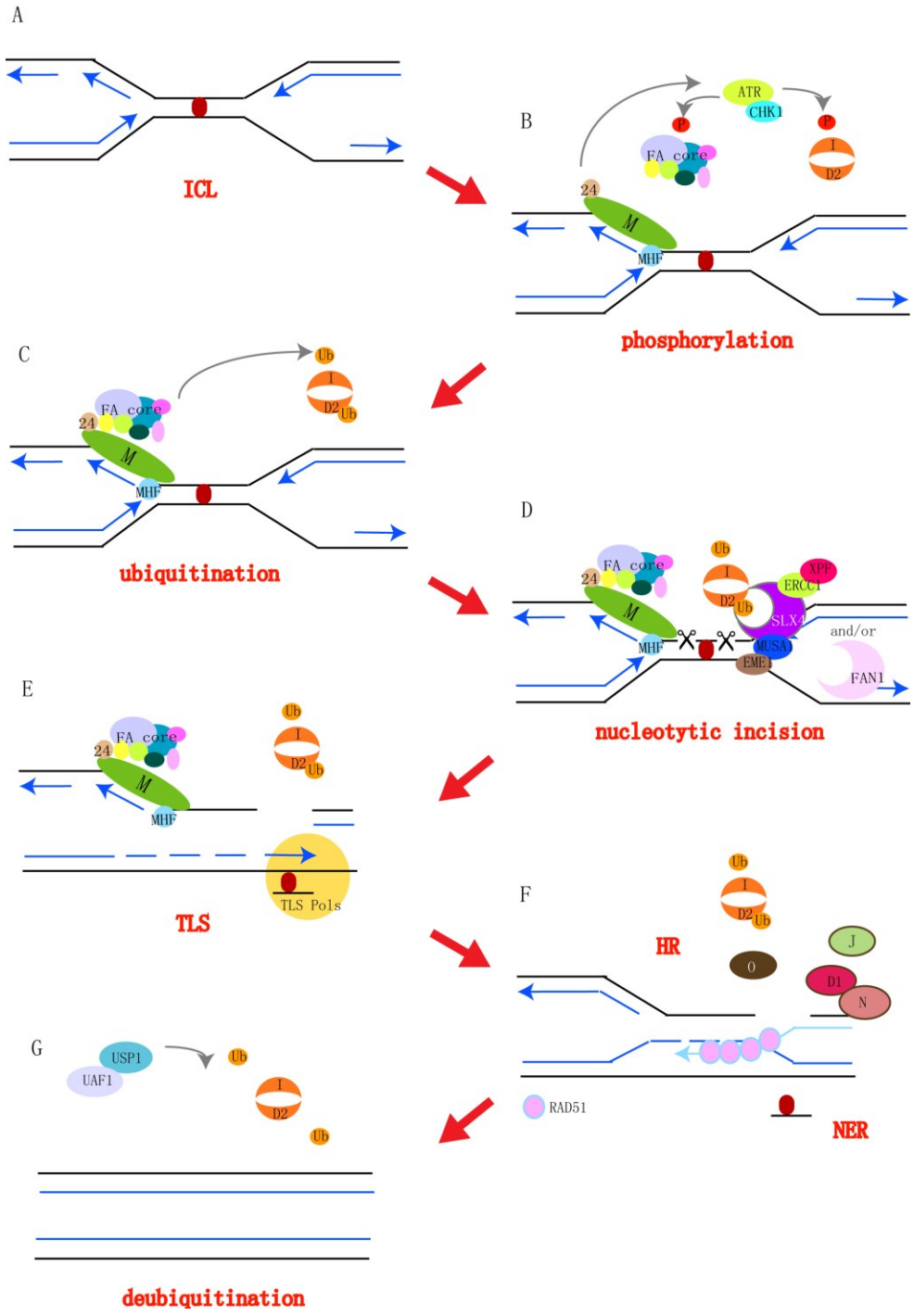

Figure 35: Regulatory role of FA proteins in ICL repair pathway

(A) DNA ICL blocks two replication forks. (B) The FANCM–FAAP24–MHF1/2 complex recognizes the stalled replication fork and recruits the FA core complex to the ICL region. FANCM also initiates ATR-CHK1-dependent checkpoint response, which phosphorylates FANCA/E/D2/I. (C) The FA core complex, a ubiquitin E3 ligase, monoubiquitinates FANCD2 and FANCI, which recruits the ID heterodimeric complex to the ICL site. (D) FANCD2-Ub acts as a platform to recruit multiple nucleases to coordinate nucleolytic incisions. FANCP/SLX4, which interacts with ERCC1–XPF and MUS81–EME1 structure-specific nucleases, and FAN1 5′-flap endonuclease are good candidates for this process. Both SLX4 and FAN1 contain UBZ4, which is essential for FANCD2-Ub-dependent recruitment to the ICL. (E) Unhooking leaves cross-linked nucleotides attached to the complementary strand, which is bypassed by TLS. (F) HR repairs the DSB that is created by incision. Downstream FA proteins promote RAD51-dependent strand invasion and the resolution of recombinant intermediates. NER removes remaining adducts and fills the gap. (G) The USP1–UAF1 DUB complex removes monoubiquitin from FANCD2-I and finishes the repair. Adapted from Kim et al. (242)

Initiation of the FA pathway occurs through formation of a heterodimer composed of FANCM and FAAP24 (FA-associated protein 24 kDa) (243), called the anchor complex. This complex recognizes the DNA lesion and recruits the FA core complex, consisting of seven FA proteins (FANCA/B/C/E/F/G/L) and two FA-associate proteins (FAAP20 and FAAP100). Aside from its function in recruitment, the FANCM-FAAP24 complex also stabilizes the stalled replication fork and initiates ATR (ataxia-telangiectasia and Rad3-related)-mediated checkpoint signaling (243, 244) . The stable association of FANCM with chromatin is maintained by the histone fold protein 1 (MHF1) and MHF2, which also amplify the activation of the pathway (245, 246). During activation, multiple FA proteins are phosphorylated by ATR-CHK1 checkpoint kinases,

representing a close interconnection of the FA pathway with DNA damage response signaling (Figure 35B).

The FA core complex, which functions as a multi-subunit ubiquitin E3 ligase complex, activates the monoubiquitination of the ID complex (FANCD2 and FANCI) after DNA damage (234). This is the key regulatory step in the pathway. The FANCL subunit of the core complex, which contains a RING domain, is an ubiquitin E3 ligase. Together with the UBE2T E2 enzyme, it conjugates a single ubiquitin to K561 and K523 of human FANCD2 and FANCI, respectively (Figure 35C) (247, 248). The monoubiquitinated ID complex is recruited to DNA lesions, where it acts as a landing pad to recruit downstream FA proteins (D1/J/N/O/P) for ICL repair activity. Initially, nucleases such as FAN1 (FA associated nuclease 1) and the newly identified FA complementation group FANCP/SLX4 protein are recruited to the monoubiquitinated FANCD2 in order to initiate nucleolytic incision (Figure 35D) (249). FAN1 and SLX4 contain a UBZ4 (ubiquitin-binding zinc finger 4) domain that can specifically recognize the ubiquitin moiety of FANCD2 (250, 251). SLX4-associated MUS81-EME1 and XPF-ERCC1 nucleases promote cross-link unhooking (252) ; this unhooking process converts a stalled replication fork into a double-strand break (DSB). Subsequently, translesion synthesis (TLS) pathway bypasses the unhooked cross-linked oligonucleotides and restores a nascent strand (Figure 35E). The DSB is then repaired by homologous recombination (HR), while nucleotide excision repair (NER) excises the remaining

adducts, and the gap is filled by DNA polymerases (Figure 35F). In the HR process, the downstream FA proteins D1/J/N/O facilitate RAD51 loading and the resolution of recombination intermediates. Finally, the modified ID complex is deubiquitinated by the deubiquitinating (DUB) enzyme USP1 (ubiquitin specific peptidase 1), associated with its activating partner, UAF1 (USP1-associated factor 1) (Figure 35G)(253, 254). The USP1-dependent deubiquitination constitutes another critical repair step in the completion of DNA ICL repair.

4.1.2 FAAP20, an integral component of the FA core complex

FAAP20 (FA-associated protein 20kD) is the newest member of the FA core complex, which was discovered very recently (255-258). FANCA/B/C/E/F/G/L are *bona fide* FA genes with hundreds of unique mutations described in the Fanconi Anemia Mutation Database (**http://www.rockefeller.edu/fanconi/**). FAAP20 is required for core complex assembly and function, but no disease-associated mutations have been found in this protein as yet (208). FAAP20 interacts with FANCA through its N-terminus (possibly through its WALLER motif) (256), maintains the integrity of the core complex, and promotes FANCD2 mono-ubiquitination (208, 256-258). The C-terminal region of FANCA (residues 1095-1200) is also required for interaction with FAAP20 (208, 256). The protein stability of FAAP20 and FANCA rely on each other. FAAP20 contains a ubiquitin binding zinc finger (UBZ) domain (255-258) of the UBZ2 type, though it has also been referred to as a UBZ4 type. The UBZ zinc finger, beginning at cysteine-147,

binds possibly preferentially to K63-linked poly-ubiquitin chains at sites of DNA damage (208, 258) and recruits monoubiquitinated REV1 in the REV1-Pol ζ complex, potentially linking the FA and translesion synthesis (TLS) pathways (198) (Figure 36). The UBZ domain of FAAP20 is necessary for normal FANCD2 monoubiquitination, but it is not involved in binding to ubiquitinated FANCD2 (60). Instead, FAAP20 may be particularly important for the correct recruitment of monoubiquitinated FANCD2 at the DNA damage site, although the importance of FAAP20 in this context appears to vary depending on which cell type is analyzed (258). The FA pathway is regulated by the ubiquitin signaling cascade initiated by RNF8 and its partner UBC13 as ubiquitin E3 and E2 enzyme, respectively, and mediated by FAAP20. FAAP20 preferentially binds the ubiquitin product of RNF8-UBC13, and this ubiquitin-binding activity and RNF8-UBC13 are both required for recruitment of FAAP20 to ICLs (258).

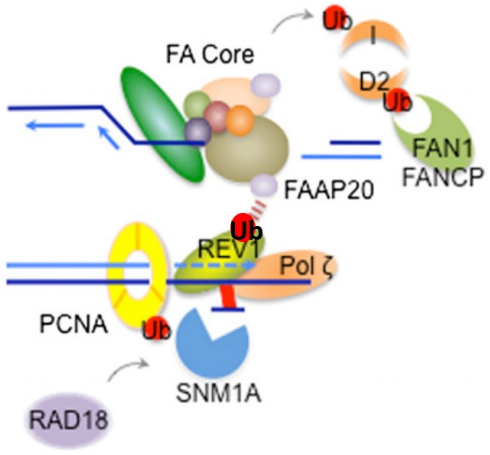

Figure 36: Regulation of TLS in replication-associated ICL repair.

FAAP20 recruits TLS machinery to ICL repair by interacting with ubiquitinated Rev1 through its UBZ domain. Adapted from Kim et al. (242)

4.1.3 Ubiquitin-binding zinc fingers (UBZs) in DNA repair proteins

Ubiquitin plays a crucial role in the regulation of DNA repair processes (79). Ubiquitin signaling is mediated by UBDs that recognize various ubiquitin modifications. Among these more than 20 different UBD classes that exist to interact with various ubiquitin modifications (32), the UBZ family is particularly interesting because it exclusively appears in proteins involved in DNA ICL repair and TLS (79). The UBZ domain was first identified by yeast two-hybrid screening and bioinformatics analysis (51), characterized by short mono-nucleate Zn-fingers within their minimal interaction region and related to a large class of DNA-binding Zn-fingers. Among the proteins identified was TAX1BP1 with two copies of a C2H2-finger (259), referred as UBZ1 (TAX1BP1-family) and the uncharacterized protein FLJ44922 (later identified as FAAP20) with one copy of a C2HC-finger, referred as UBZ2 (FLJ44922/FAAP20 family). A bioinformatics search for Zn-fingers with more similarity to UBZ1/2 than to the DNA-binding Zn fingers revealed a number of additional UBZ candidate families, UBZ3 through UBZ9.

Nowadays, there are two subfamilies of UBZ domains in DNA repair pathways: (1) UBZ3, which contains a highly conserved C2H2 zinc finger and (2) UBZ4 which is defined as a RAD18-like C2HC zinc finger (Figure 37). So far, the only UBZ3 domain was identified in yeast Rad30 and its mammalian homolog, Pol η. The NMR structure of the Pol η UBZ3 domain was determined (68), revealing a similar fold as the classic

DNA-binding zinc finger, which contains two short β strands and a α helix in which the ubiquitin binding surface is mapped on the α-helix on the opposite side of the zinc-coordinating residues by NMR titration and mutagenesis experiments. In contrast, the UBZ4 domain has been found in many proteins to be involved in DNA repair, such as Pol κ, WRNIP1 (Werner helicase-interacting protein 1), and RAD18 (79) . Pol κ interacts with ubiquitinated PCNA, while WRNIP and RAD18 bind to K48- and K63-poly-ubiquitin chains (201, 224, 260, 261). Additionally, many of these proteins are involved in ICL repair. For example, SNM1A, a mammalian ortholog of yeast Pso2, was shown to target a stalled replication fork by recognizing PCNA-Ub via its UBZ4 domain (262). Besides SNM1A, the recently identified FAN1 nuclease and the FANCP/SLX4 nuclease scaffold protein also contain UBZ4 domains that specifically recognize FANCD2-Ub and function in the downstream ICL repair process, emphasizing a specialized role of UBZ4-containing proteins in mediating ICL repair (250, 263). So far, only two of the UBZ4 domain structures are determined, WRNIP and RAD18 in complex with monoubiquitin (67). They both contain a ββα fold, and interact with ubiquitin in an almost identical mode (usage of β1 and α-helix), but completely different from the predicted UBZ3-ubiquitin interaction mode. In summary, UBZ domains are found in diverse players in the DNA repair signaling pathways and represent a key signature of DNA repair proteins that interact with ubiquitinated targets or polyubiquitin chains at DNA damage sites to fulfill their functions.

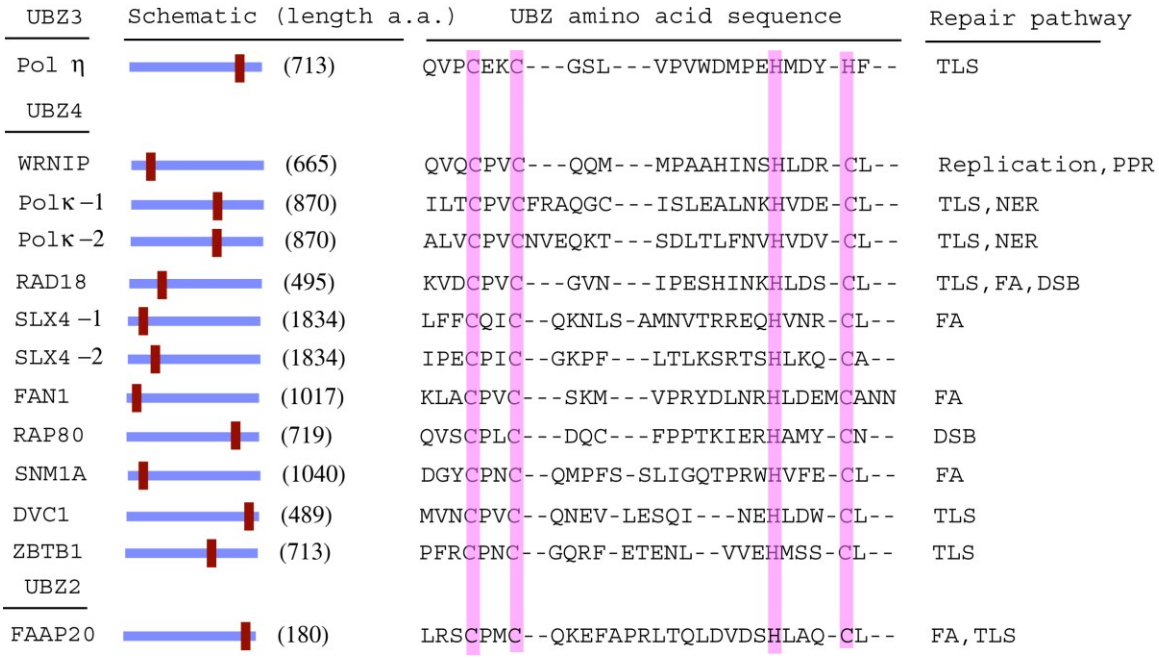

Figure 37: UBZ domains in DNA repair proteins

Various amino acid sequences of known UBZ domains are shown, and their relative positions are marked as a red bar in the blue protein schematic. The conserved residues that comprise a zinc-binding core are shaded in the pink box.

4.1.4 The FAAP20 UBZ domain

The UBZ domain of the human FAAP20 protein is located at the C-terminus, from residues 144 to 173. Based on primary sequence alignment, it contains a CCHC motif to coordinate with the zinc ion. It is predicted to have two β strands and one α-helix to adopt the canonical zinc-finger ββα fold. The FAAP20 UBZ is conserved in higher vertebrates, but not in yeast (Figure 38).

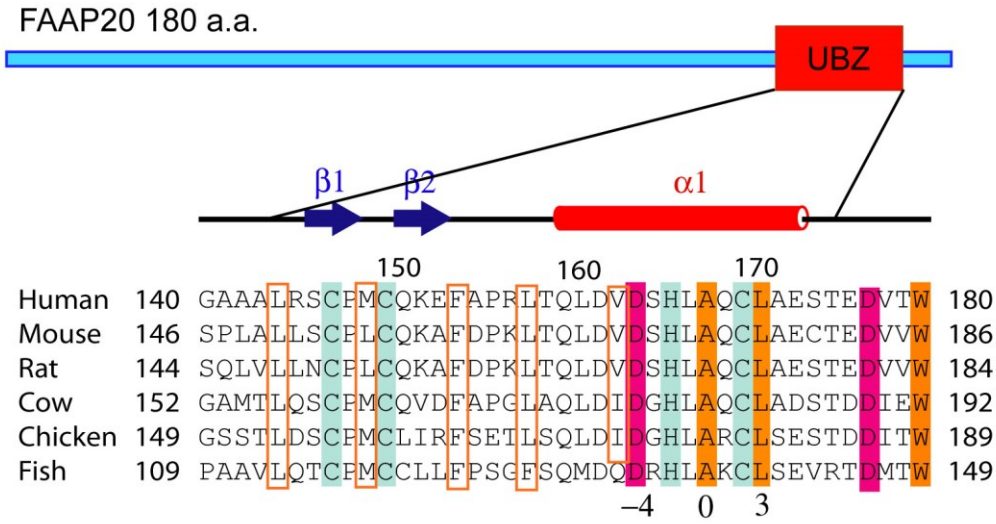

Figure 38: Sequence alignment of the FAAP20 UBZ from different species.

The conserved zinc-coordinating residues are highlighted in blue. Conserved residues involved in the formation of the hydrophobic core of the UBZ domain are boxed in orange and those involved in ubiquitin interaction are highlighted. Acidic residues are colored in pink and hydrophobic residues are colored in orange.

The interaction of FAAP20 UBZ with ubiquitin has been probed by *in vitro* pulldown experiments. *E.coli* expressing WT FAAP20 can pulldown ubiquitin chains, while the FAAP20 containing a mutation of the zinc-coordinating cysteine to alanine cannot. However, the exact species of ubiquitin that interacts with FAAP20 UBZ remains controversial. For monoubiquitin, *E.coli* expressed His-Flag-FAAP20 cannot pulldown monoubiquitin *in vitro*. (255) A recombinant GST-UBZ domain also cannot pulldown monoubiquitin *in vitro* (258). However, recombinant Ubi-GST can pulldown either MBP-FAAP20 (256) or Flag-tagged FAAP20 (257) *in vitro*. With regard to the ubiquitin chain specificity, *E.coli* expressed His-Flag-FAAP20 can only pulldown K63-linked ubiquitin

chains, but not K48-linked ubiquitin chains (255). *E.coli* expressed GST-FAAP20 can pulldown both K48-linked and K63-linked ubiquitin chains from 293T cell lysates (256). A recombinant GST-UBZ domain can pulldown both K48-linked and K63-linked ubiquitin chains with preference toward K63-linked chains (258). Since our NMR titration experiments (below) clearly shows the interaction between FAAP20 UBZ and monoubiquitin, the discrepancy of these pulldown experiments likely reflects the different experimental conditions.

FAAP20 UBZ is required for DNA-damage-induced chromatin loading of FANCA (255), for normal recruitment of the FA core complex and FANCD2 to ICL (258) and for bridging the interaction of FA complex with the translesion synthesis machinery (255, 258). In particular, the FAAP20 UBZ-ubiquitin interaction has been shown to be essential for recruiting the monoubiquitinated Rev1(257) (Figure 36).

4.2 Results

4.2.1 Solution structure of the human FAAP20 UBZ domain

In order to probe the molecular basis of ubiquitin recognition by the FAAP20 UBZ, we first determined the solution structure of the apo protein, consisting of human FAAP20 residues 140-180, by NMR (Figure 39; statistics shown in Table 3). The solution structure of the FAAP20 UBZ domain was determined with 852 NOE, 44 dihedral angle, and 28 hydrogen bond constraints. Residues 144-173 of FAAP20 UBZ domain consists of two short antiparallel b-strands formed by residues 145-147 and 151-153, respectively,

and a C-terminal a-helix spanning residues 160-173, adopting a well-converged canonical ββα fold, with mean pairwise r.m.s. deviations of 0.29 and 1.11 Å for the backbone and heavy atoms, respectively. The ββα fold is held together by zinc coordination of conserved "finger" residues, including C147 from β1, C150 from the 'fingertip' of the β1-β2 loop, and H166 and C170 from the α-helix. A C147A or C150A mutation in the human FAAP20 UBZ domain resulted in loss of capability of interacting with ubiquitin, increased sensitivity to ICL agents, and inactivation of FA pathway. Packing of the two β-strands against the α-helix is augmented by interactions among conserved hydrophobic residues, including L144 N-terminal to β1, M149 at the fingertip of the β1-β2 loop, F154 C-terminal to β2, L158 N-terminal to the α-helix, and V163 and L167 in the middle of the α-helix (Figure 38). Outside of the compact ββα zinc-finger module, N-terminal residues 140-143 and C-terminal residues 174-180 are completely disordered.

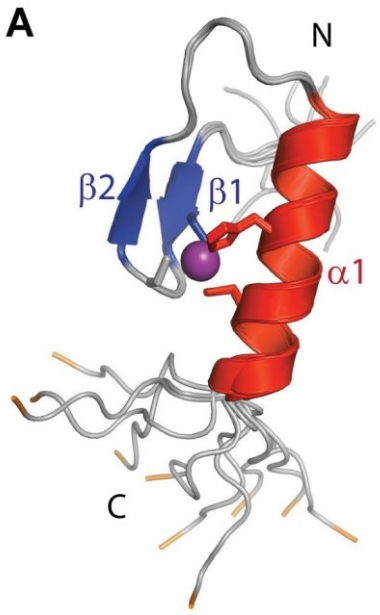

Figure 39: Solution structure of the human FAAP20 UBZ

FAAP20 UBZ adopts a canonical ββα fold with the β-strands shown in blue, and α-helix shown in red. Sidechains of the zinc-coordinating CCHC motif are shown in the stick model and colored according to the secondary structures. The zinc atom is shown in the sphere model and colored in purple. The backbone of the terminal tryptophan, W180, is colored in orange.

Table 3: Structural statistics of human FAAP20 UBZ domain

Structural statistics for the human FAAP20 UBZ (10 structures)[a]	
FAAP20 UBZ (140-180)	
NOE distance restraints	852
Intra-residue	144
Sequential	246
Medium-range ($1< \|i-j\| \leq 4$)	278
Long-range ($\|i-j\| \geq 5$)	184
Hydrogen bond constraints[b]	28
Dihedral angle constraints[c]	44
Target function value	1.04 ± 0.01
Ramachandran plot[d]	
Favored region (98%)	91.9
Allowed region (>99.8%)	100.0
Mean pairwise RMSD (FAAP20 UBZ 144-173)	
Backbone	0.29 ± 0.07 Å
Heavy Atoms	1.11 ± 0.18 Å

[a] None of these structures exhibit distance violations greater than 0.5 Å or dihedral angle violations greater than 5°.
[b] Two constraints ($d_{HN-O} \leq 2.5$ Å and $d_{N-O} \leq 3.5$ Å) are used for each identified hydrogen bond.
[c] Dihedral angle constraints were generated by talos+ based on backbone atom chemical shifts (108), and by analysis of NOE patterns.
[d] MolProbity was used to assess the quality of the structures (219).

4.2.2 The disordered C-terminal tail of FAAP20 UBZ is involved in ubiquitin binding

After determining the apo structure of human FAAP20 UBZ domain, we went on to probe the FAAP20-ubiquitin interaction by NMR titration, in which we observed extensive resonance perturbation for a lot of residues. Surprisingly, we noticed large

chemical shift perturbation for the residues of the disordered C-terminal tail in addition to residues of the ββα UBZ module. The backbone resonance of the terminal Trp residue (W180) of FAAP20, in particular, undergoes a prominent change of chemical shift upon ubiquitin binding (Figure 40 left panel), indicating that it is experiencing a major alteration of the chemical environment and may likely be directly involved in ubiquitin binding.

In order to further examine the potential involvement of the disordered C-terminal tail in ubiquitin binding, we conducted the ^{1}H-^{15}N heteronuclear NOE experiment, looking for changes in flexibility of the FAAP20 protein backbone upon ubiquitin binding (Figure 40 right panel). In this experiment, a negative value of the heteronuclear NOE reflects rapid conformational fluctuation at the ps-to-ns timescale that is typically found in disordered loops, whereas a positive value reflects a lack of fast motions that is consistent with a rigid conformation commonly found in well-folded proteins (215). In the apo state, residues of the C-terminal tail of FAAP20 UBZ all displayed negative heteronuclear NOEs, consistent with the disordered conformation observed in the NMR ensemble of apo FAAP20. In contrast, these values became distinctly positive upon addition of 2-molar excess of ubiquitin, indicating that the C-terminal tail has gained rigidity upon the formation of the FAAP20-ubiquitin complex. Taken together, these data strongly support the involvement of the C-terminal tail of FAAP20, in addition to the canonical UBZ domain, in interaction with ubiquitin *in vitro*.

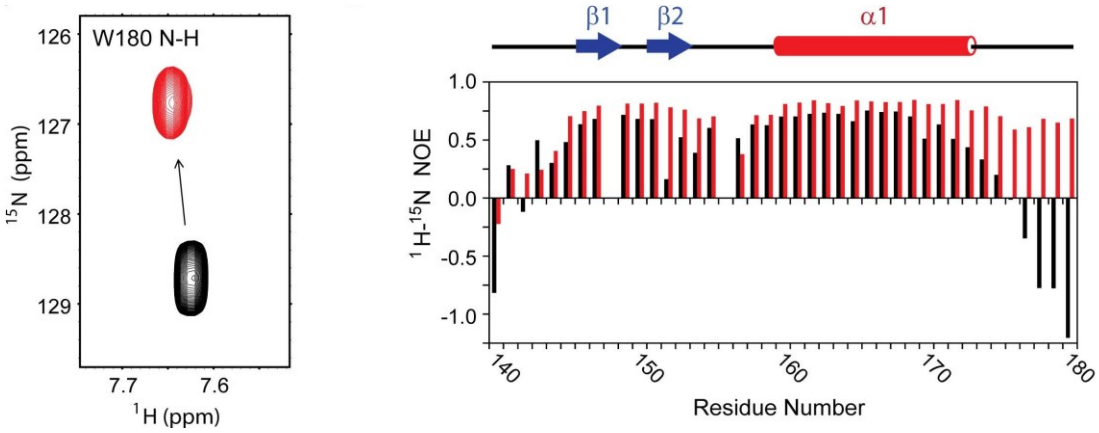

Figure 40: C-terminal tail of human FAAP29 UBZ is involved in ubiquitin binding

(left) Titration of unlabeled ubiquitin into ^{15}N-labeled FAAP20 significantly perturbs the backbone amide resonance of W180. Apo FAAP20 is shown in black and FAAP20 with excess ubiquitin in red. (right) ^{1}H-^{15}N heteronuclear NOEs plotted per residue of FAAP20. Values for the apo FAAP20 are shown with black bars and values for the FAAP20 with excess ubiquitin are shown with red bars.

4.2.3 Solution structure of the human FAAP20 UBZ-ubiquitin complex

The unexpected involvement of the disordered C-terminal tail outside the FAAP20 UBZ for ubiquitin interaction is distinct from all other UBZ domains studied thus far (68, 69, 264, 265) and warrants a thorough structural investigation. Using sparsely-sampled 3-D and 4-D NMR spectroscopy (266), we have determined the solution structure of the FAAP20 UBZ-ubiquitin complex (Figure 41AB), with the ensemble mean pairwise r.m.s. deviations of 0.55 and 1.16 Å for the backbone and heavy atoms, respectively (Table 4).

Binding of FAAP20 causes little conformational change for ubiquitin, which adopts the familiar α/β-roll topology with the main α-helix wrapped in a distorted

central β-sheet consisting of five strands (Figure 41B). At the opposite, solvent-exposed surface of the ubiquitin β-sheet lies a cluster of hydrophobic residues centered around $I44^{Ub}$ for high-affinity interaction with FAAP20 (Figure 41C). In contrast, binding of FAAP20 to ubiquitin is accompanied by a major conformational switch, involving binding-induced attachment of the disordered C-terminal tail of apo FAAP20 to its compact ββα zinc finger module to form an expanded ubiquitin binding interface (compare Figure 39 with Figure 41BD).

The overall assembly of the FAAP20 UBZ-ubiquitin complex is similar to that of the MIU/IUIM-ubiquitin complex (41, 42) and to the predicted structural model of the Pol η UBZ-ubiquitin complex (68), with the prominent UBZ helix packing against the solvent-exposed surface of the central β-sheet of ubiquitin. The FAAP20 UBZ helix is oriented in parallel with the central β-strand (β5) of ubiquitin, and the C-terminal end of the UBZ helix is located in close proximity to the C-terminus of ubiquitin (Figure 39C). At the center of the UBZ helix lies an invariant Ala residue (A168) that wedges into the hydrophobic pocket of ubiquitin encircled by $L8^{Ub}$, $I44^{Ub}$ and $V70^{Ub}$. Along the same face of the FAAP20 helix, at one helical turn N-terminal to the central A168 (at the -4 position in the primary sequence, Figure 38) lies a conserved Asp residue (D164) that forms hydrogen bonds with the backbone of $A46^{Ub}$ and $G47^{Ub}$, anchoring the N-terminal half of the UBZ helix to ubiquitin. Such a helix-ubiquitin interaction is augmented by hydrophobic interactions at the C-terminal half of the UBZ helix, with the invariant

hydrophobic residue L171 located at one helical turn C-terminal to A168 (at the +3 position, Figure 38) along the same face of the UBZ helix to interact with $I44^{Ub}$, $V70^{Ub}$, and the side chain of $R42^{Ub}$, further enhancing the binding toward ubiquitin. A group of less conserved residues along the UBZ helix also contribute to the human FAAP20 interaction with ubiquitin, including S165 that forms a hydrogen bond with $H68^{Ub}$, the sidechain of Q169 that interacts with $L8^{Ub}$, and A172 that interacts with $V70^{Ub}$ (Figure 41D).

Unique to the FAAP20 UBZ-ubiquitin interface is the presence of binding-induced folding of a disordered C-terminal tail of FAAP20 containing two invariant residues: Asp (D177) and Trp (W180) (Figure 38). Although residues of this tail are highly dynamic in the apo state as reflected by their negative heteronuclear NOE values (Figure 41D), they form an extended β-loop that is fixated by joint interactions with ubiquitin and with the core module of the FAAP20 UBZ. In particular, the sidechain of D177 points toward the sidechains of $R42^{Ub}$ and $R72^{Ub}$, potentially latching onto ubiquitin through salt bridges with these residues. W180, the Trp residue at the very C-terminus of this tail in FAAP20, appears to play a crucial role in mediating FAAP20-ubiquitin binding. Emanating toward $I44^{Ub}$ from the now extended C-terminal loop of FAAP20 that stretches along the UBZ-ubiquitin interface, the indole group of W180 is affixed to the carbonyl group of $G47^{Ub}$ located within the β3-β4 loop of ubiquitin through a hydrogen bond of the W180 imino group (Figure 41E). Such an interaction is

supported by the observed intermolecular NOE between W180 aromatic protons and the Hα proton of $K48^{Ub}$ (Figure 41F, lower right panel). The W180 indole ring points toward $I44^{Ub}$ and is juxtaposed between the sidechains of $K48^{Ub}$ and $Q49^{Ub}$ of the β3-β4 loop of ubiquitin, and P148 and M149 of the fingertip and L167 of the central helix of the FAAP20 UBZ. Importantly, W180, together with a cluster of conserved residues of the compact zinc finger module of FAAP20, including L167, A168, and L171, forms an encircled hydrophobic pocket around $I44^{Ub}$ for high-affinity ubiquitin binding (Figure 41CD). Accordingly, numerous intermolecular NOEs are observed in the 4-D omit NOE (difference) spectrum between $I44^{Ub}$ and surrounding FAAP20 residues (Figure 41F and Figure 42). Highlighting the integral structure of the FAAP20 C-terminal tail in ubiquitin recognition, residues of this tail all display positive heteronuclear NOEs that are congruent with FAAP20 residues of the ββα zinc-finger module (Figure 41D).

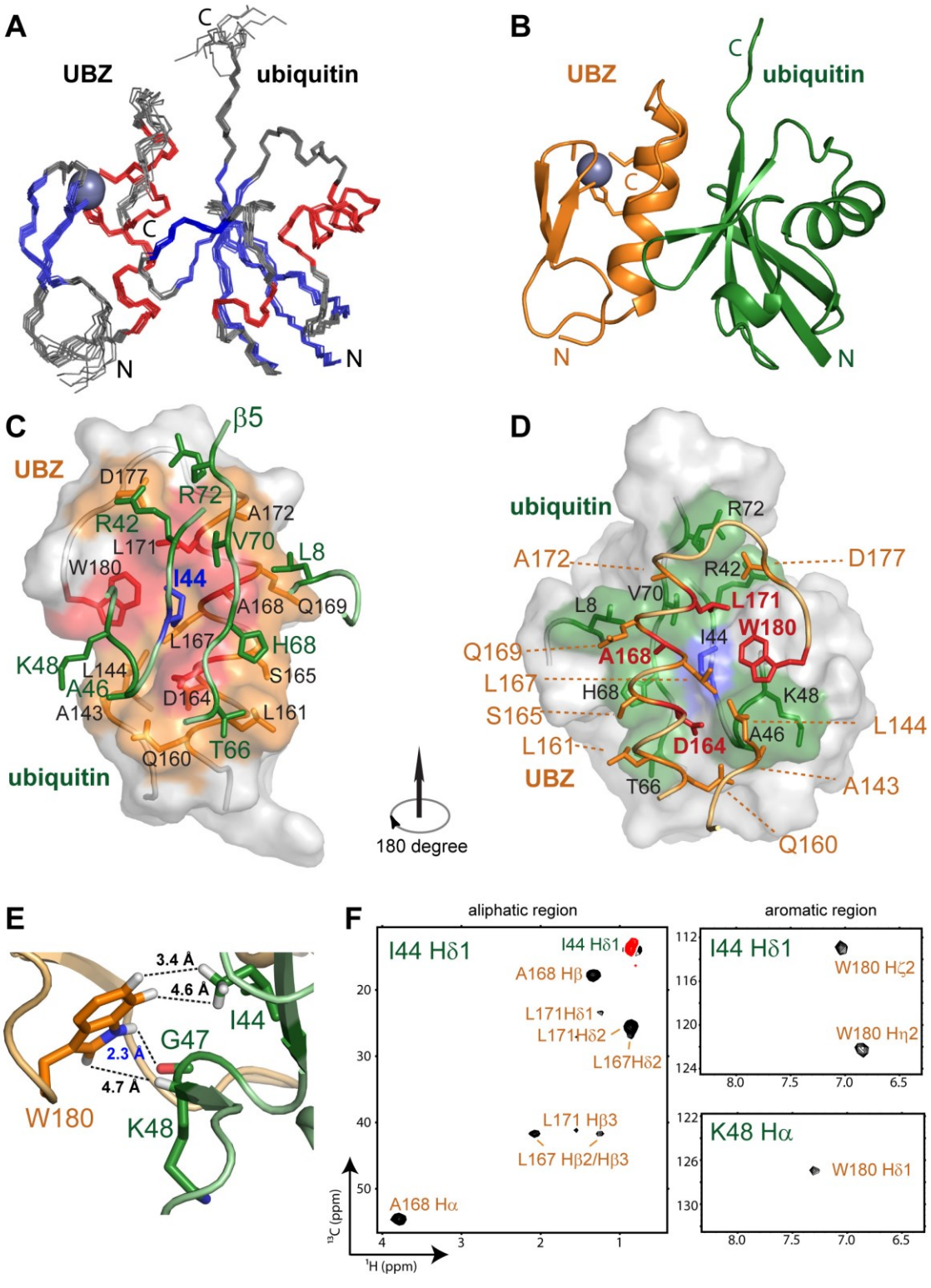
A
UBZ
ubiquitin
C
N
B
UBZ
ubiquitin
C
N
C
UBZ
β5
D177
R72
A172
R42
V70
L8
L171
W180
I44
A168
Q169
L167
H68
K48
L144
S165
A46
D164
A143
L161
T66
Q160
ubiquitin
180 degree
D
ubiquitin
R72
A172
D177
R42
V70
L8
L171
W180
I44
Q169
A168
H68
K48
L167
S165
L144
A46
L161
D164
UBZ
T66
A143
Q160
E
3.4 Å
4.6 Å
I44
G47
2.3 Å
4.7 Å
W180
K48
F
aliphatic region
aromatic region
I44 Hδ1
I44 Hδ1
A168 Hβ
L171Hδ1
L171Hδ2
L167Hδ2
L171 Hβ3
L167 Hβ2/Hβ3
A168 Hα
W180 Hζ2
W180 Hη2
K48 Hα
W180 Hδ1
13C (ppm)
1H (ppm)

Figure 41: Solution structure of the human FAAP20-ubiquitin complex.

(A) Backbone traces of the NMR ensemble of 10 structures. β-strands, α-helices and loops are colored in blue, red and grey, respectively. Zinc is shown as a grey sphere. (B) Ribbon diagram of the complex, with FAAP20 in orange and ubiquitin in green. (C) The surface of FAAP20 UBZ with the interface is colored in orange. The D-A-L motif and the C-terminal tryptophan residue are colored in red. The sidechains of the interfacial residues are shown in the stick model and labeled in black. The ubiquitin is colored in green with the sidechains of the interfacial residues shown in the stick model. The central residue $I44^{Ub}$ is colored in blue. (D) The surface of ubiquitin with the interface is colored in green. The central residue $I44^{Ub}$ is colored in blue. The sidechains of the interfacial residues are shown in the stick model and labeled in black. The FAAP20 UBZ is colored in orange with the sidechains of the interfacial residues shown in the stick model. The D-A-L motif and the C-terminal tryptophan residue are colored in red. The two β-strands and the zinc ion are not shown for clarity. (E) Interaction between FAAP20 W180 (orange) and ubiquitin I44, G47 and K48 (green). The interacting sidechains are shown in the stick model. The hydrogen bond between W180 Hε1 and G47 backbone is shown as a dashed line and the distances are labeled in blue. The hydrophobic interactions supported by intermolecular NOEs are shown as dashed lines and the distances are labeled in black. (F) Representative 2-D slices of the 4-D ^{13}C-HMQC-NOESY-HSQC omit spectra centered on I44 Hδ1 and K48 Hα, respectively, illustrating the intermolecular NOEs from FAAP20 to ubiquitin I44 Hδ1 and K48 Hα protons. Positive intermolecular NOE crosspeaks are colored in black and a negative diagonal peak is colored in red. Resonances of FAAP20 and ubiquitin are colored in orange and green, respectively. The generation of omit spectra is shown in Figure 42 in detail.

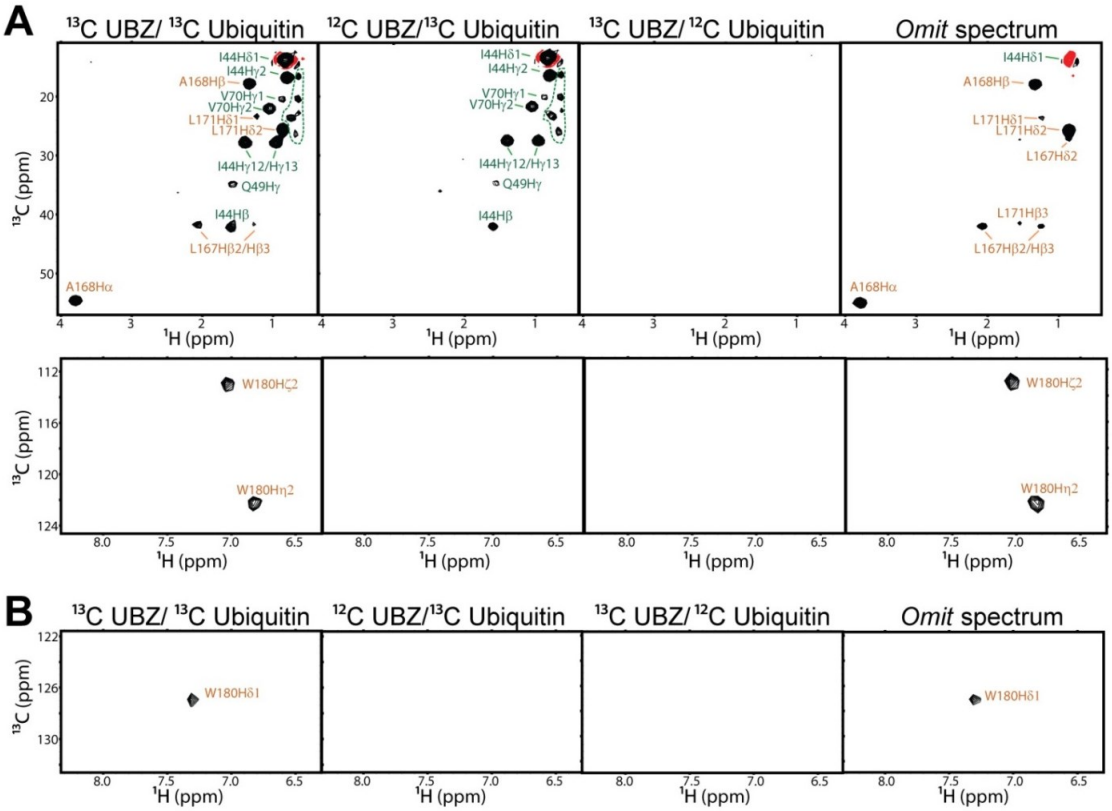

Figure 42: Intermolecular NOE difference (omit) spectrum of the human FAAP20-ubiquitin complex

Sparsely-sampled 4-D 13C-HMQC–NOESY–HSQC spectra were collected for the FAAP20-ubiquitin complex with both components or with individual components 13C-labeled as described previously (3). Reconstruction of the difference time domain signals of the uniformly labeled protein complex from component-labeled samples generated an omit spectrum containing only intermolecular NOEs. Slight over-subtraction of time domain data from individual components generates negative diagonal signals (red) in the omit spectrum and ensures that all of the positive crosspeaks originate from intermolecular NOEs. Panel (A) (upper, aliphatic regions; lower, aromatic regions) shows sections of F1-F2 slices of the corresponding 4-D spectra centered at 13.52 ppm in F3 and 0.796 ppm in F4, displaying NOEs to $I44^{Ub}$ Hδ1. Peaks circled by dash lines are off-plane signals. Panel (B) shows sections of F1-F2 slices of the corresponding 4-D spectra centered at 53.88 ppm in F3 and 4.764 ppm in F4 in the aromatic region, displaying NOEs to the ubiquitin K48 Hα.

Table 4: Structual statistics for the human FAAP20 UBZ-ubiquitin complex

Structural statistics for the human FAAP20 UBZ-ubiquitin complex (10 structures)[a]	
FAAP20 UBZ (140-180)	
NOE distance restraints	1028
Intra-residue	126
Sequential	291
Medium-range (1< |i-j| ≤4)	309
Long-range (|i-j| ≥ 5)	302
Hydrogen bond constraints[b]	28
Dihedral angle constraints[c]	55
Ubiquitin (1-76)	
NOE distance restraints	2867
Intra-residue	328
Sequential	616
Medium-range (1< |i-j| ≤4)	569
Long-range (|i-j| ≥ 5)	1354
Hydrogen bond constraints[b]	60
Dihedral angle constraints[c]	132
Intermolecular NOE distance constraints	187
Target function value	2.38 ± 0.09
Ramachandran plot[d]	
Favored region (98%)	90.8
Allowed region (>99.8%)	99.3
Mean pairwise RMSD	
(FAAP20 UBZ 143-180; Ubiquitin 1-74)	
Backbone	0.55 ± 0.08 Å
Heavy Atoms	1.16 ± 0.08 Å

[a] None of these structures exhibit distance violations greater than 0.5 Å or dihedral angle violations greater than 5°.

[b] Two constraints ($d_{HN-O} \leq 2.5$ Å and $d_{N-O} \leq 3.5$ Å) are used for each identified hydrogen bond.

[c] Dihedral angle constraints were generated by talos+ based on backbone atom chemical shifts (108), and by analysis of NOE patterns.

[d] MolProbity was used to assess the quality of the structures (219).

4.2.4 The conserved D-A-L motif of the FAAP20 UBZ helix contributes to ubiquitin binding

After determining the solution structure of the FAAP20 UBZ-ubiquitin complex, we investigated the FAAP20-ubiquitin interaction by isothermal titration calorimetry (ITC) measurements, probing the contribution of individual residues to the binding affinity. Titration of wild-type (WT) ubiquitin into WT FAAP20 UBZ revealed a K_d value of 9.26 μM (Table 5), a binding affinity stronger than the affinities of most ubiquitin binding domain (UBD)-ubiquitin interactions.

Table 5: Binding affinities of the human FAAP20 UBZ-ubiquitin complexes measured by ITC

FAAP20 UBZ	ubiquitin	K_d (μM)
WT	WT	9.26
WT	I44A	NDB
D164A	WT	NDB
A168Y	WT	NDB
L171A	WT	529
W180A	WT	NDB

NDB: No detectable binding or too weak to fit reliably (Kd > 600 μM)

Among the interface residues of the FAAP20-ubiquitin complex, $I44^{Ub}$ is at the center of the FAAP20 UBZ-ubiquitin interaction and displays numerous intermolecular NOEs to FAAP20 residues, including W180 of the C-terminal tail (Figure 41F and Figure 42). Unsurprisingly, mutation of $I44^{Ub}$ to Ala completely abolished the ubiquitin binding by FAAP20, verifying that ubiquitin is being recognized through the canonical hydrophobic patch centered at $I44^{Ub}$ (Table 5, Figure 43). On the FAAP20 side, A168 is located at the center of the UBZ helix and plays a pivotal role in anchoring the UBZ helix

to the conserved hydrophobic pocket of ubiquitin formed by $L8^{Ub}$, $I44^{Ub}$, and $V70^{Ub}$. Accordingly, substitution of A168 by Tyr eliminated the FAAP20-ubiquitin interaction. Point mutation of D164A of the FAAP20 UBZ similarly disrupted ubiquitin binding in our ITC studies (Table 5, Figure 43), consistent with a previous report (255). Finally, mutation of L171A significantly weakened the FAAP20 interaction with ubiquitin (Table 5, Figure 43). Taken together, these observations further support the structurally observed binding mode of the FAAP20-ubiquitin complex, corroborating the recognition mode of the D-A-L motif along the surface of the UBZ α-helix for ubiquitin interaction.

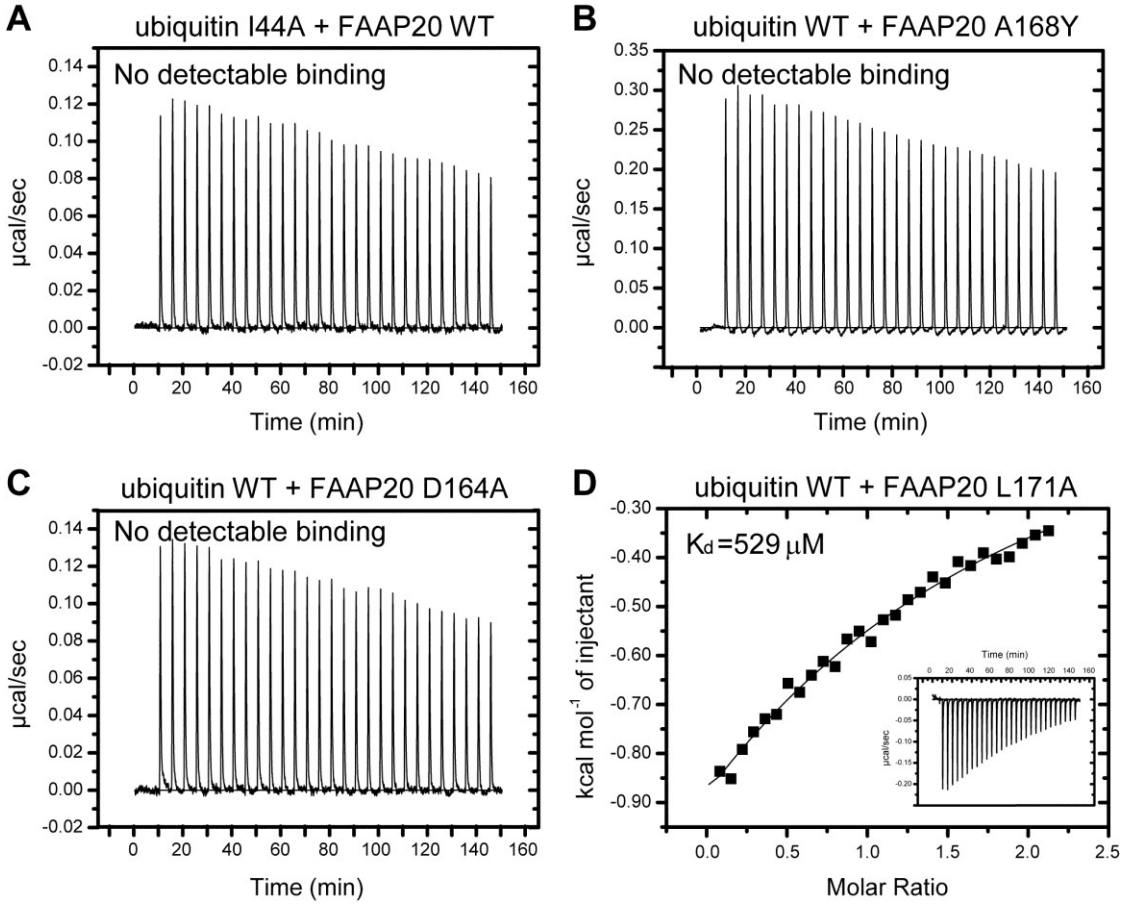

Figure 43: Mutations of the interface residues affect the FAAP20-ubiquitin binding.

Raw data of heat change per injection and the fitted affinity curve, whenever applicable, are shown for (A) the ubiquitin I44A mutant, (B) the FAAP20 A168Y mutant, (C) the FAAP20 D164A mutant, and (D) the FAAP20 L171A mutant, respectively.

4.2.5 The terminal tryptophan is required for FAAP20-ubiquitin binding *in vitro* and efficient ICL repair *in vivo*

Since the FAAP20-ubiquitin interaction uniquely features an expanded ubiquitin-binding interface beyond the compact ββα zinc-finger module, with the absolutely conserved Trp residue (W180) of the disordered C-terminal tail of the apo FAAP20 participating in numerous interactions with ubiquitin residues in the protein

complex, including $I44^{Ub}$ at the center of the ubiquitin interface, we conducted *in vitro* and *in vivo* experiments to evaluate the consequence of a W180A mutation in affecting the FAAP20-ubiquitin interaction and in FA core complex mediated ICL repair. In order to make sure that the W180A mutation does not alter the UBZ structure, a ^{1}H-^{15}N HSQC spectrum of the FAAP20 W180A mutant was collected. With the exception of the few C-terminal residues neighboring W180 in the primary sequence, the overall spectra of the WT FAAP20 UBZ and the W180A mutant are nearly superimposable, consistent with the notion of a disordered C-terminal tail in the apo FAAP20 and verifying that the W180A mutation does not disturb the ββα fold of the UBZ (Figure 44). Surprisingly, in contrast to the tight binding of the WT FAAP20 toward ubiquitin (Figure 45A), the W180A mutant completely abolished the FAAP20 interaction with ubiquitin in ITC measurements (Figure 45B), revealing that W180 not only contributes to, but is also required for ubiquitin binding *in vitro*.

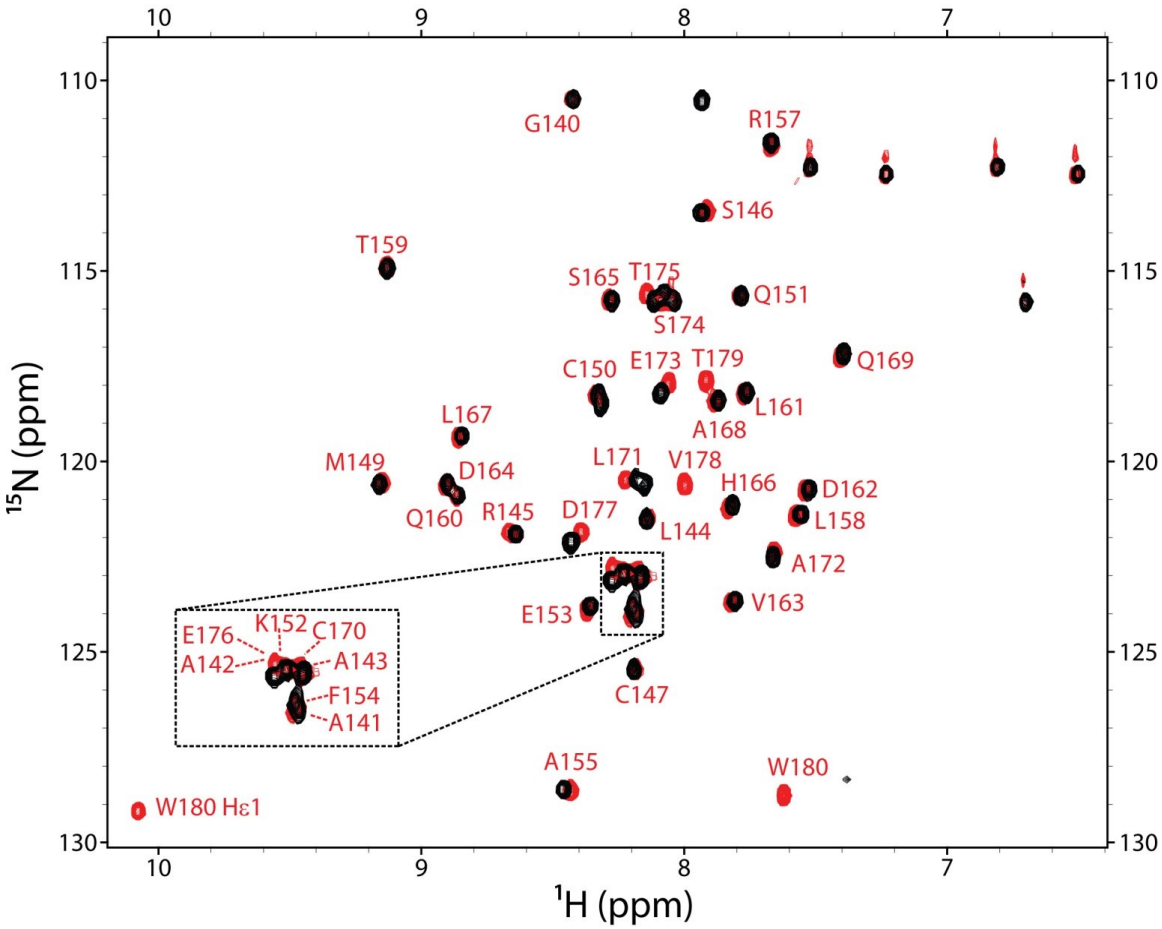

Figure 44: ^{1}H-^{15}N HSQC spectra of the human FAAP20 UBZ WT protein (red) and W180A mutant (black).

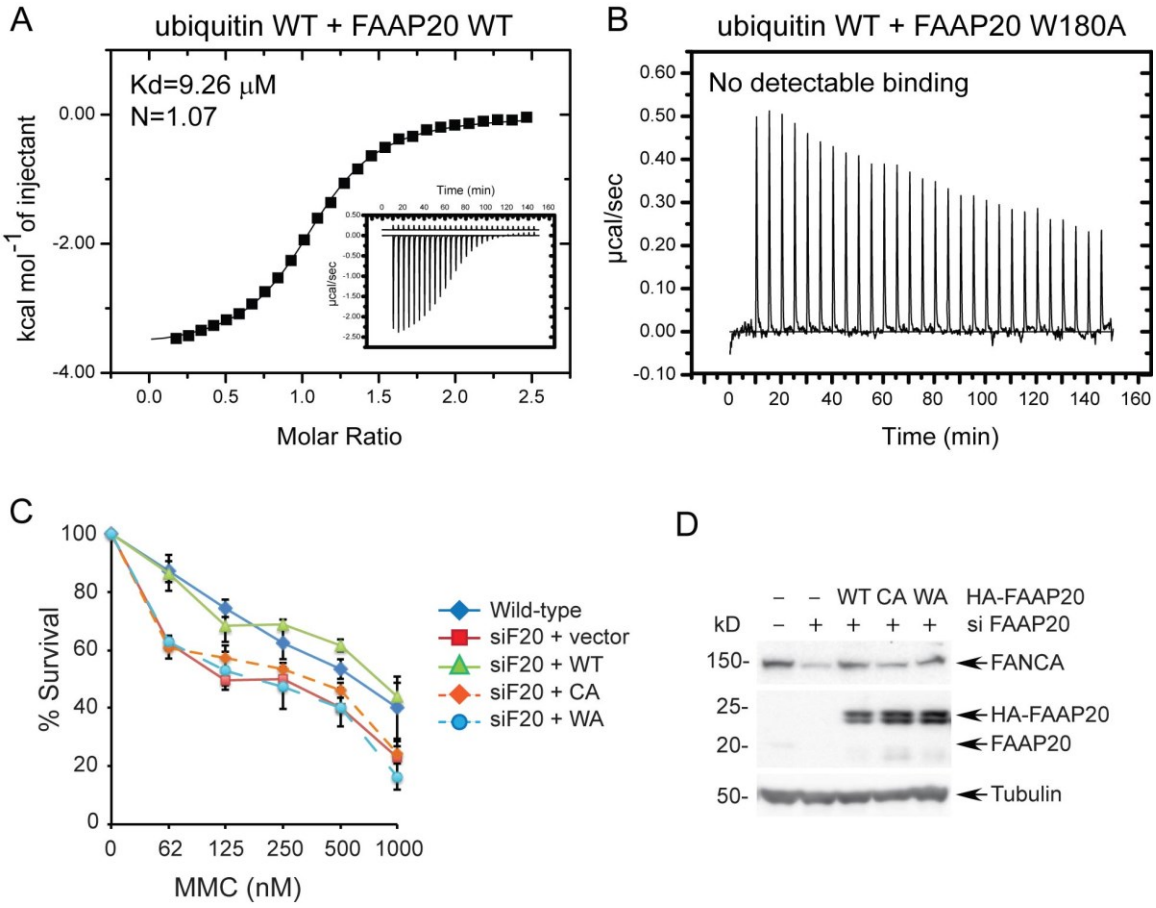

Figure 45: The C-terminal tryptophan residue (W180) of FAAP20 outside the compact UBZ module plays an indispensable role in ubiquitin-recognition of FAAP20 in vitro and in efficient ICL DNA repair in vivo.

(A) Measurement of binding affinity between the WT FAAP20 and WT ubiquitin by ITC. (B) Mutation of W180A in FAAP20 abolished ubiquitin binding in ITC measurements. (C) U2OS cells stably expressing siRNA-resistant FAAP20 wild-type, C147A & C150A (CA), or W180A (WA) were transfected with siRNA against FAAP20 (siF20) for 48 h, and cell viability was determined 6 days after the treatment of indicated doses of mitomycin C. (D) Immunoblot analysis of U2OS cells in (C) harvested 72 h after siRNA treatment.

In order to assess the contribution of the C-terminal tryptophan residue in regulating DNA repair *in vivo*, we also tested the ability of full-length FAAP20 harboring a W180A mutation to complement the DNA damage sensitivity of the FAAP20-depleted

mammalian cells. The integrity of the FAAP20 UBZ motif is required for conferring cellular resistance to DNA ICL-inducing agent, mitomycin C (MMC) (255). Therefore, we replaced endogenous FAAP20 depleted by siRNA with siRNA-resistant FAAP20 variants and challenged cells with MMC. Stable expression of siRNA-resistant WT FAAP20 could restore the hypersensitivity to MMC caused by FAAP20 depletion, while neither C147A & C150A (the mutations disrupting the zinc finger fold) nor the W180A mutant could rescue the DNA damage sensitization phenotype (Figure 45CD). Such an observation further emphasizes the functional importance of W180 and its C-terminal tail residing outside the core UBZ domain in mediating the FAAP20 interaction with ubiquitin, which is essential for the ICL DNA repair *in vivo*. It is intriguing to note that FAAP20 has 8 isoforms, but only isoforms 2 and 6 contain an intact UBZ domain with a C-terminal Trp residue and hence are capable of interaction with ubiquitin and execution of ICL DNA repair. The functions of the other isoforms, however, await to be determined.

4.3 Discussion

4.3.1 Distinct ubiquitin recognition by human FAAP20 UBZ

Depending on the organization of the zinc coordinating residues and ubiquitin binding motifs in the primary sequence, UBZ domains are classified into different subgroups, including the UBZ3 type of zinc fingers consisting of CCHH zinc-coordinating residues represented by the Pol η UBZ, and the UBZ4 type of zinc fingers

consisting of CCHC zinc-coordinating residues that are found in a variety of proteins involved in DNA repair activities. Structural comparison of the FAAP20 UBZ-ubiquitin complex to the WRNIP UBZ-ubiquitin complex (PDB 3VHT), which shares a similar ubiquitin-binding interface with the recently reported Rad18 UBZ-ubiquitin complex (67), reveals a completely different orientation of the central UBZ helix, even though the FAAP20 UBZ is sometimes referred to as a UBZ4 domain (198, 256), the same type as the UBZs of WRNIP and Rad18 (Figure 46AB). Further structural comparison shows ubiquitin recognition by the FAAP20 UBZ helix is actually similar to that of the MIU/IUIM helix in the ubiquitin-bound complex (41, 42), a binding mode that features a D(-4)-A(0)-L(+3) motif along the ubiquitin-recognition α-helix and is also shared by the Pol η UBZ, a UBZ3 type of zinc finger (68) (Figure 46CD). The structural observation of a distinct ubiquitin binding mode of the FAAP20 UBZ from that of a UBZ4 type zinc finger lends further credence to the suggestion of the FAAP20 UBZ as the founding member of the UBZ2 family that is characterized by the CCHC zinc-coordinating residues and the DxHxAxCL motif (79).

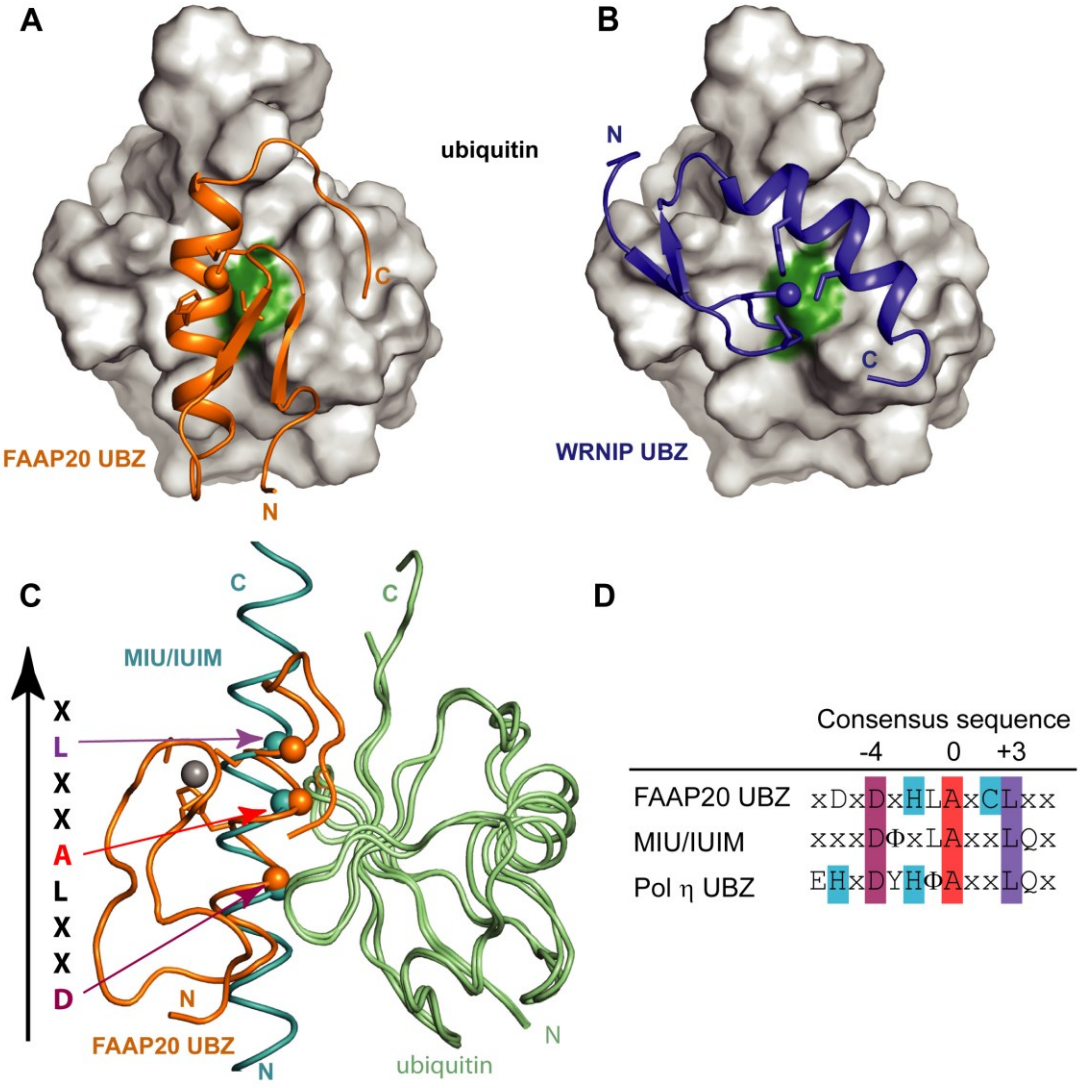

Figure 46: Ubiquitin recognition by the human FAAP20 UBZ helix is distinct from that of the WRNIP UBZ, but show similarity to that of IUIM/MIU.

Panels (A) and (B) show the ribbon diagrams of the FAAP20-ubiquitin complex (orange, this work) and the WRNIP UBZ-ubiquitin complex (blue, PDB 3VHT), respectively. Ubiquitin is shown in identical orientation and colored in grey surface with central interfacial residue I44 in green. (C) Overlay of the human FAAP20 UBZ (orange)-ubiquitin (pale green) complex (this work) with IUIM/MIU (cyan)-ubiquitin (pale green) complex (PDB 2FIF). The conserved interfacial residues along the α-helix are shown in sphere and indicated by arrows colored according to the sequence alignment in (D). (D)

Alignment of the consensus sequences of the FAAP20 UBZ domain with MIU/IUIM and Pol η UBZ domain. The central invariant alanine is highlighted in red, the conserved aspartate at the -4 position in magenta, the conserved leucine at the +3 position in purple and the zinc ligands in blue.

4.3.2 Binding-induced folding of the C-terminal tail of the human FAAP20 UBZ domain

All of the UBZ fingers that have been characterized thus far interact with ubiquitin exclusively through the compact ββα zinc finger module. FAAP20 also contains a well-defined UBZ, together with a C-terminal tail that is disordered in the absence of ubiquitin. Unexpectedly, the structural elucidation of the FAAP20-ubiquitin complex reveals a binding-induced folding of the disordered C-terminal tail of FAAP20 that expands the ubiquitin-binding interface beyond the canonical UBZ module for high-affinity interaction with ubiquitin. In particular, the very C-terminal Trp residue (W180), upon transforming the disordered tail of FAAP20 into an extended β-loop, buries its sidechain into the interface of the FAAP20 UBZ and ubiquitin to fortify the FAAP20-ubiquitin interaction. The FAAP20 mutant with alanine substitution of the terminal Trp residue (W180A) abolishes ubiquitin binding *in vitro* and ICL DNA repair activity *in vivo*, emphasizing the crucial contribution of the FAAP20 C-terminal tail to FAAP20 function and highlighting a unique binding mode not observed in any of the previously characterized UBZ-ubiquitin complexes.

4.4 Materials and Methods

4.4.1 Molecular cloning

The DNA sequence of residues 140-180 of human FAAP20 was synthesized; the PCR-amplified DNA was double digested and ligated into a modified pET15 vector (EMD Biosciences, Inc.) between the NdeI and XhoI restriction sites. The final construct contained an N-terminal His_{10} tag, followed by a GB1 solubility enhancement tag (225). A TEV cleavage site is engineered between the GB1 tag and the FAAP20 sequence. The ubiquitin construct is the same as used in Chapter 3. The presence of the correct inserts of these constructs was confirmed by DNA sequencing.

4.4.2 Protein purification

The purification of ubiquitin is the same as described in Chapter 2. The His_{10}-GB1-fused FAAP20 construct was overexpressed in *Escherichia coli* BL21 (DE3*) cells (Invitrogen). Bacterial cells were cultured in M9 minimal media using ^{15}N-NH_4Cl and ^{13}C-glucose as the sole nitrogen and carbon sources (Cambridge Isotope Laboratories), and induced by IPTG (0.1 mM IPTG at 20 °C for 18 hr). 50 μM $ZnSO_4$ was added to the cell culture at the time of induction. The overexpressed proteins were purified by a Ni^{2+}-NTA column; then the N-terminal His_{10}-GB1 tag of the FAAP20 construct were removed by TEV (homemade) cleavage, respectively. Another Ni^{2+}-NTA column was used to remove the His_{10}-GB1-tag and the His_6-TEV protease from the UBZ. The UBZ protein was further purified by size-exclusion chromatography (Superdex 75, GE Healthcare).

4.4.3 NMR sample preparation and data collection

NMR samples of the apo FAAP20 UBZ and the FAAP20 UBZ-ubiquitin complex were exchanged into an NMR buffer containing 25 mM sodium phosphate, 100 mM KCl, and 10% D_2O or 100% D_2O (pH=7.0) and concentrated to a final protein concentration of 3 mM of FAAP20 alone or with 1:1 molar ratio of FAAP20 and ubiquitin, respectively.

NMR experiments were conducted at 25 °C using Agilent INOVA 600 or 800 MHz spectrometers. Four pairs of sparse-sampled 3-D triple-resonance experiments were collected for backbone resonances assignment. 4-D sparsely-sampled HC(co)NH-TOCSY and HCCH-TOCSY experiments were collected for the side-chain resonances assignment. The traditional 3-D ^{15}N-separated NOESY-HSQC experiments and the 4-D sparsely-sampled ^{13}C-HMQC-NOESY-^{15}N-HSQC and ^{13}C-HMQC-NOESY-HSQC experiments using a uniformly ^{15}N, ^{13}C-labeled sample were collected for NOE assignments.

NMR data were processed by NMRPIPE (151) and SCRUB (141). NMR spectra were analyzed with XEASY (153) and SPARKY (152).

4.4.4 Generation of 4-D omit spectra and identification of intermolecular NOEs

For intermolecular NOE identification, three identical 4-D ^{13}C HMQC-NOESY-HSQC experiments were recorded for each sample (uniformly ^{15}N, ^{13}C-labeled, ubiquitin selectively labeled and UBZ selectively labeled). Each experiment used a mixing time of 150ms, with spectral widths of 70ppm for the ^{13}C dimensions (t1 and t3) and 10ppm for

the ^{1}H(t2) dimension and t_{max} of 4.60ms for the ^{13}C (t1 and t3) dimensions and 11.70ms for the ^{1}H (t2) dimension. Each experiment was recorded for 86 hr. The omit spectrum was generated by subtracting scaled FIDs of UBZ-labeled complexes and ubiquitin-labeled complex from those of the uniformly-labeled complex with the scaling factor of 1.05 and 1.17, respectively and reconstructed by SCRUB (141). Intermolecular NOE crosspeaks from the omit spectrum were analyzed manually.

4.4.5 Structure calculation

For apo FAAP20, NOEs identified from 3-D ^{15}N-separated NOESY-HSQC, 4-D sparsely-sampled ^{13}C-HMQC-NOESY-^{15}N-HSQC and ^{13}C-HMQC-NOESY-HSQC experiments were used for automatic CYANA (149) structure calculation and refinement in the presence of dihedral angle constraints derived from TALOS+ (108) analysis of chemical shift information. The final structural ensembles (10 structures) of apo FAAP20 display no NOE violations > 0.5 Å and no dihedral angle violations > 5 °. The statistics of the structural ensemble is shown in Table 3. The apo FAAP20 structure has been deposited in RCSB and BMRB (PDB ID: 2muq; BMRB ID: 25229).

For the FAAP20-ubiquitin complex, intermolecular NOEs were identified from the 4-D omit spectra as described previously. Intermolecular NOE crosspeaks were analyzed manually and converted into distance constraints using the calibration module in CYANA (149). NOE crosspeaks from 3-D ^{15}N-separated NOESY-HSQC and 4-D sparsely-sampled ^{13}C-HMQC-NOESY-^{15}N-HSQC and ^{13}C-HMQC-NOESY-HSQC using a

uniformly labeled complex sample were used for automatic CYANA (149) structure calculation and refinement in the presence of manually assigned intermolecular NOE constraints and dihedral angle constraints derived from TALOS+ (108) analysis of chemical shift information. The final structural ensembles (10 structures) of the FAAP20 UBZ-ubiquitin complex display no NOE violations > 0.5 Å and no dihedral angle violations > 5 °. The statistics of the structural ensemble is shown in Table 4. The FAAP20-ubiquitin complex structure has been deposited in RCSB and BMRB (PDB ID: 2mur; BMRB ID: 25230).

4.4.6 Isothermal titration calorimetry

The wild-type or mutant human ubiquitin sample (final concentrations in the 2–3 mM range) was titrated into a solution of FAAP20 (wild-type and mutants, 0.2–0.3 mM range) in a buffer containing 25 mM sodium phosphate, 100 mM KCl, pH 7.0. Twenty-eight injections of 10 μl each were performed at 25 °C using a VP-ITC Microcalorimeter (GE Healthcare), and data were analyzed using the Origin software assuming one-site binding (Origin Lab).

4.4.7 Heteronuclear NOE experiment

The samples for the heteronuclear ^{1}H-^{15}N NOE (215) experiment were prepared by mixing 1mM uniformly ^{15}N-labeled FAAP20 with non-labeled ubiquitin in two different ratios, 1:0 (apo) and 1:2, using the same buffer condition in the NMR structural study. The ^{1}H-^{15}N NOE values were calculated from the ratios of peak intensities in the

saturated versus unsaturated spectra. This work was done by Qinglin Wu (Zhou Lab, Duke University)

4.4.8 Cell culture, plasmid construction and transfection, siRNA

U2OS and 293T cells were cultured in Dulbecco's Modified Eagles Medium supplemented with 10% fetal bovine serum (FBS) following standard culture conditions and procedures. Generation of FAAP20 constructs was described previously (257). Point mutations were introduced by QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) and confirmed by DNA sequencing. Stable U2OS cells were generated by retroviral transduction of siRNA-resistant pMSCV-Flag-HA-FAAP20 variants followed by 2 μg/mL puromycin selection.

Plasmid transfection for retroviral transduction was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocols. siRNA duplexes were synthesized by Qiagen and transfected using Lipofectamine RNAiMAX (Invitrogen). The targeting sequence for FAAP20 is 5′-CACGGTGAGCCCGGAGCTGAT, and the nucleotides changed in the siRNA-resistant construct are shown in lower cases, 5′-gACtGTtAGtCCtGAaCTaAT.

4.4.9 Protein analysis and antibodies

Cells were lysed with NETN300 buffer (300 mM NaCl, 0.2 mM EDTA, 50 mM Tris [pH 7.5], 1 % NP40) supplemented with protease inhibitor cocktail (Roche). Cellular lysates were resolved by NuPAGE (Invitrogen) gels and transferred onto PVDF

membrane (EMD Millipore) followed by immunoblotting using antibodies as indicated: anti-FANCA (Bethyl), anti-FAAP20 (Sigma Atlas), and anti-Tubulin (Sigma). Signals were detected by either enhanced chemiluminescence method (Western Lightening, Perkin Elmer) or LAS-4000 Imaging system (GE Healthcare Life Sciences).

4.4.10 Cytotoxicity assay

siRNA-treated U2OS cells were seeded on 96-well plates and treated with increasing doses of mitomycin C (Sigma) the following day. Cell viability was determined using the Cell Titer-Glo Luminescence Cell Viability Assay kit (Promega) and Spectramax M5 (Molecular Devices) 6 days following continued drug treatment.

5. Conclusions and future Studies

5.1 Conclusions

Protein-protein interactions play important roles in almost every aspect of the cellular processes, by regulating the function of signaling pathways or assembling the protein machinery. Structural analysis of protein complexes can provide atomic information about the functions and regulatory mechanisms of the protein complexes. Additionally, the structure information of protein complexes can be used for developing novel therapeutics, transforming basic science knowledge to clinical drugs.

The advancement of NMR technology has greatly enhanced the sensitivity of NMR experiments and has made it increasingly feasible to determine high-resolution structures of protein complexes that play important roles in various cellular processes by solution NMR. In comparison with the well-established procedure to determine the high-resolution structure of a single-chain protein, structure determination of a protein complex presents unique challenges. In particular, the limited number of interfacial NOEs that define macromolecular complexes require reliable detection and unambiguous assignment to ensure the proper assembly of individual components of the protein complex. The recent development of fast acquisition of high-resolution 4-D NOESY experiments based on sparse time-domain sampling and high-performance spectral reconstruction in the frequency domain offers a convenient solution to this challenge. In this dissertation, I present a new strategy for reliable detection and

assignment of intermolecular NOEs by 4-D difference NMR experiments for structure determination of protein complexes. This strategy can generate high-sensitivity, high-resolution 4-D "omit" spectrum containing only intermolecular NOEs. It overcomes the incomplete suppression of intramolecular NOEs in the half-filtered experiment, and thus guarantees the cleanness of the omit spectrum. Each NOE crosspeak in the omit spectrum can be identified based on two chemical shift parameters: proton and the attached carbon attached to the proton, which dramatically reduces the assignment ambiguity.

The impact of such a strategy was first demonstrated in automated structure determination of the homooligomer foldon complex. We show that intermolecular NOEs derived from the 4-D omit spectrum can be used for automated structural analysis of the foldon complex, and the calculated structure overlay perfectly with the crystal structure. In contrast, the automated structural analysis using intermonomer NOE information extracted from traditional 3-D half-filtered experiment failed to converge, most likely due to the interference of incompletely suppressed intramolecular NOEs and signal degeneration in the indirect ^{1}H dimension.

By applying this new strategy and other 4-D fast NMR based experiments, we next studied the molecular mechanism of ubiquitin recognition in DNA damage response by determining two ubiquitin-UBD complex structures: human Pol ι UBM1-ubiquitin complex and human FAAP20 UBZ-ubiquitin complex. Due to the α-helical

property and large portion of loop region in the UBM1 domain, spectrum overlapping hindered accurate and reliable data analysis with conventional 3-D experiments. The usage of 4-D fast NMR techniques resolved this problem to a great extent, giving better signal separation and clean detection of NOE crosspeaks. The human Pol ι UBM1-ubiquitin complex structure was determined at high-resolution, revealing a novel helix-turn-helix fold of UBM1 for interacting with the solvent-exposed β-sheet of ubiquitin, centering on L8 instead of the canonical I44 of ubiquitin.

Also by using 4-D fast NMR and 4-D omit NOE techniques, we determined the human FAAP20 UBZ-ubiquitin complex structure, which is a critical linkage between the Fanconi anemia pathway and TLS that fulfills the function of DNA interstrand crosslink repair. FAAP 20 contains a classical ββα zinc-finger module with conserved CCHC motif coordinating the zinc ion. We showed that the FAAP20-ubiquitin interaction extends beyond the canonical UBZ domain and is accompanied by transforming the disordered C-terminal tail of FAAP20 into a rigid β-loop, with the invariant C-terminal tryptophan (W180 of human FAAP20) emanating toward $I44^{Ub}$ for enhanced binding. From the 4-D omit spectrum, we detected several confident intermolecular NOEs from W180 of FAAP20 with I44 and K48 of ubiquitin, supporting interaction between the FAAP20 C-terminal tail and ubiquitin. This binding-induced folding of the C-terminal tail outside the UBD domain reveals a novel ubiquitin recognition mode, which likely contributes to ubiquitin-chain specificity and to the

higher binding affinity between ubiquitin and UBDs. Again, the use of 4-D omit NOE spectroscopy has enabled unambiguous detection of numerous intermolecular NOEs, including the C-terminal tryptophan residue, of FAAP20 toward ubiquitin.

5.2 Future directions

The small sizes of UBDs make them ideal systems for structural analysis of their ubiquitin-bound complexes by solution NMR. It will be interesting to apply the 4-D fast NMR technique to larger protein complex systems. Similar 4-D noesy experiments based on fast NMR have been successfully applied to structural studies of several proteins in our lab, including Ssu72 (141, 267), HCAII (141) and MBP (268), proving the feasibility of applying 4-D noesy experiments to protein sizes of up to 42kDa. The 4-D omit spectrum present in this thesis has better signal-to-noise ratio compared to normal 4-D C noesy spectrum. This is because 1) the strongest diagonal signals have been removed which reduces the strongest aliasing artifacts that comes with the diagonal signals; and 2) the intramolecular NOEs have been removed, which dramatically reduces the total number of peaks in the spectrum, also resulting many fewer aliasing artifacts. Therefore, it is feasible to apply the 4-D omit NOE spectrum strategy to larger protein complexes, which can take advantage of better signal separation of 4-D fast NMR to facilitate data analysis of large protein complex systems. Membrane protein complexes will be another important field for the application of 4-D fast NMR. The structural discrepancy of the DAGK trimeric complex clearly highlights such a need. By

applying advanced 4-D fast NMR techniques to this system, we might be able to detect and assign key intersubunit NOEs to resolve the controversy of the NMR structure. Although the requirement of detergent micelle for membrane proteins increases the molecular weights of the protein-micelle complexes dramatically, which may negatively impact the application of fast NMR studies due to rapid signal relaxation, recent demonstration of fast NMR experiments of large membrane proteins in detergent micelles and nanodiscs suggests the feasibility of detailed structural analysis of DAGK by 4-D fast NMR experiments, including omit NOE experiments described here.

Ubiquitin recognition plays an important regulation role in DNA damage response pathways. One critical feature of ubiquitination is the diversity of the ubiquitin chain due to different types of possible linkages and different numbers of ubiquitin moieties within a chain. Our studies of the UBM and UBZ domains have so far been restricted to monoubiquitin interactions, though there is a growing body of evidence to suggest that many DNA repair proteins are capable of recognizing polyubiquitin chains, in particular K63-linked polyubiquitin chains that have been implicated as being recognized by both UBM and UBZ. Whether these domains are capable of recognizing K63 ubiquitin chains independently or in cooperation with another ubiquitin binding domain remains to be elucidated. In particular, our FAAP20 UBZ-ubiquitin complex structure shows the utilization of the C-terminal tail of FAAP20 for ubiquitin recognition that extends the interface to encompass K48 of ubiquitin. This suggests the possible

structural basis for linkage type specificity of FAAP20-ubiquitin interaction because the K48 linkage might be blocked by the interaction between the C-terminal tail of FAAP20 and ubiquitin. Extended structural studies of FAAP20 UBZ in complex with K48 or K63-linked diubiquitin will provide evidence for this hypothesis. These studies will shed new insights into the regulation of DNA repair pathways, and they will undoubtedly benefit from novel NMR methodology such as fast NMR and high-resolution 4-D omit NOE spectroscopy as described in this dissertation.

References

1. Wagner MJ, Stacey MM, Liu BA, & Pawson T (2013) Molecular mechanisms of SH2- and PTB-domain-containing proteins in receptor tyrosine kinase signaling. *Cold Spring Harbor perspectives in biology* 5(12):a008987.

2. Finley D & Chau V (1991) Ubiquitination. *Annual review of cell biology* 7:25-69.

3. Goldstein G, *et al.* (1975) Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proceedings of the National Academy of Sciences of the United States of America* 72(1):11-15.

4. Miller J & Gordon C (2005) The regulation of proteasome degradation by multi-ubiquitin chain binding proteins. *FEBS letters* 579(15):3224-3230.

5. Elsasser S & Finley D (2005) Delivery of ubiquitinated substrates to protein-unfolding machines. *Nature cell biology* 7(8):742-749.

6. Hicke L (2001) A new ticket for entry into budding vesicles-ubiquitin. *Cell* 106(5):527-530.

7. Raiborg C, Rusten TE, & Stenmark H (2003) Protein sorting into multivesicular endosomes. *Current opinion in cell biology* 15(4):446-455.

8. Staub O & Rotin D (2006) Role of ubiquitylation in cellular membrane transport. *Physiological reviews* 86(2):669-707.

9. Huang TT & D'Andrea AD (2006) Regulation of DNA repair by ubiquitylation. *Nature reviews. Molecular cell biology* 7(5):323-334.

10. Di Fiore PP, Polo S, & Hofmann K (2003) When ubiquitin meets ubiquitin receptors: a signalling connection. *Nature reviews. Molecular cell biology* 4(6):491-497.

11. Haglund K & Dikic I (2005) Ubiquitylation and cell signaling. *The EMBO journal* 24(19):3353-3359.

12. Vijay-Kumar S, Bugg CE, Wilkinson KD, & Cook WJ (1985) Three-dimensional structure of ubiquitin at 2.8 A resolution. *Proceedings of the National Academy of Sciences of the United States of America* 82(11):3582-3585.

13. Vijay-Kumar S, Bugg CE, & Cook WJ (1987) Structure of ubiquitin refined at 1.8 A resolution. *Journal of molecular biology* 194(3):531-544.

14. Hershko A & Ciechanover A (1998) The ubiquitin system. *Annual review of biochemistry* 67:425-479.

15. Pickart CM (2001) Mechanisms underlying ubiquitination. *Annual review of biochemistry* 70:503-533.

16. Hicke L, Schubert HL, & Hill CP (2005) Ubiquitin-binding domains. *Nature reviews. Molecular cell biology* 6(8):610-621.

17. Nijman SM, *et al.* (2005) A genomic and functional inventory of deubiquitinating enzymes. *Cell* 123(5):773-786.

18. Sowa ME, Bennett EJ, Gygi SP, & Harper JW (2009) Defining the human deubiquitinating enzyme interaction landscape. *Cell* 138(2):389-403.

19. Wong E & Cuervo AM (2010) Integration of clearance mechanisms: the proteasome and autophagy. *Cold Spring Harbor perspectives in biology* 2(12):a006734.

20. Weissman AM (2001) Themes and variations on ubiquitylation. *Nature reviews. Molecular cell biology* 2(3):169-178.

21. Pickart CM & Fushman D (2004) Polyubiquitin chains: polymeric protein signals. *Current opinion in chemical biology* 8(6):610-616.

22. Welchman RL, Gordon C, & Mayer RJ (2005) Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nature reviews. Molecular cell biology* 6(8):599-609.

23. Hicke L (2001) Protein regulation by monoubiquitin. *Nature reviews. Molecular cell biology* 2(3):195-201.

24. Pickart CM (1997) Targeting of substrates to the 26S proteasome. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 11(13):1055-1066.

25. Al-Hakim A, *et al.* (2010) The ubiquitous role of ubiquitin in the DNA damage response. *DNA repair* 9(12):1229-1240.

26. Kulathu Y & Komander D (2012) Atypical ubiquitylation - the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nature reviews. Molecular cell biology* 13(8):508-523.

27. Kirisako T, *et al.* (2006) A ubiquitin ligase complex assembles linear polyubiquitin chains. *The EMBO journal* 25(20):4877-4887.

28. Tokunaga F, *et al.* (2009) Involvement of linear polyubiquitylation of NEMO in NF-kappaB activation. *Nature cell biology* 11(2):123-132.

29. Iwai K & Tokunaga F (2009) Linear polyubiquitination: a new regulator of NF-kappaB activation. *EMBO reports* 10(7):706-713.

30. Bremm A & Komander D (2011) Emerging roles for Lys11-linked polyubiquitin in cellular regulation. *Trends in biochemical sciences* 36(7):355-363.

31. Glauser L, Sonnay S, Stafa K, & Moore DJ (2011) Parkin promotes the ubiquitination and degradation of the mitochondrial fusion factor mitofusin 1. *Journal of neurochemistry* 118(4):636-645.

32. Dikic I, Wakatsuki S, & Walters KJ (2009) Ubiquitin-binding domains - from structures to functions. *Nature reviews. Molecular cell biology* 10(10):659-671.

33. Hurley JH, Lee S, & Prag G (2006) Ubiquitin-binding domains. *The Biochemical journal* 399(3):361-372.

34. Wu H, Lo YC, & Lin SC (2010) Recent advances in polyubiquitin chain recognition. *F1000 biology reports* 2:1-5.

35. Lakomek NA, *et al.* (2008) Residual dipolar couplings as a tool to study molecular recognition of ubiquitin. *Biochemical Society transactions* 36(Pt 6):1433-1437.

36. Lange OF, *et al.* (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320(5882):1471-1475.

37. Beal RE, Toscano-Cantaffa D, Young P, Rechsteiner M, & Pickart CM (1998) The hydrophobic effect contributes to polyubiquitin chain recognition. *Biochemistry* 37(9):2925-2934.

38. Mueller TD & Feigon J (2002) Solution structures of UBA domains reveal a conserved hydrophobic surface for protein-protein interactions. *Journal of molecular biology* 319(5):1243-1255.

39. Ryu KS, *et al.* (2003) Binding surface mapping of intra- and interdomain interactions among hHR23B, ubiquitin, and polyubiquitin binding site 2 of S5a. *The Journal of biological chemistry* 278(38):36621-36627.

40. Wang Q, Young P, & Walters KJ (2005) Structure of S5a bound to monoubiquitin provides a model for polyubiquitin recognition. *Journal of molecular biology* 348(3):727-739.

41. Penengo L, *et al.* (2006) Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell* 124(6):1183-1195.

42. Lee S, *et al.* (2006) Structural basis for ubiquitin recognition and autoubiquitination by Rabex-5. *Nat Struct Mol Biol* 13(3):264-271.

43. Hirano S, *et al.* (2006) Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *Nat Struct Mol Biol* 13(3):272-277.

44. Hofmann K & Bucher P (1996) The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends in biochemical sciences* 21(5):172-173.

45. Varadan R, Assfalg M, Raasi S, Pickart C, & Fushman D (2005) Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *Molecular cell* 18(6):687-698.

46. Kang RS, *et al.* (2003) Solution structure of a CUE-ubiquitin complex reveals a conserved mode of ubiquitin binding. *Cell* 113(5):621-630.

47. Akutsu M, *et al.* (2005) Structural basis for recognition of ubiquitinated cargo by Tom1-GAT domain. *FEBS letters* 579(24):5385-5391.

48. Kawasaki M, *et al.* (2005) Molecular mechanism of ubiquitin recognition by GGA3 GAT domain. *Genes to cells : devoted to molecular & cellular mechanisms* 10(7):639-654.

49. Lo YC, *et al.* (2009) Structural basis for recognition of diubiquitins by NEMO. *Molecular cell* 33(5):602-615.

50. Rahighi S, *et al.* (2009) Specific recognition of linear ubiquitin chains by NEMO is important for NF-kappaB activation. *Cell* 136(6):1098-1109.

51. Bienko M, *et al.* (2005) Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* 310(5755):1821-1824.

52. Alam SL, *et al.* (2004) Ubiquitin interactions of NZF zinc fingers. *The EMBO journal* 23(7):1411-1421.

53. Reyes-Turcu FE, *et al.* (2006) The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell* 124(6):1197-1208.

54. Pai MT, *et al.* (2007) Solution structure of the Ubp-M BUZ domain, a highly specific protein module that recognizes the C-terminal tail of free ubiquitin. *Journal of molecular biology* 370(2):290-302.

55. Brzovic PS, Lissounov A, Christensen DE, Hoyt DW, & Klevit RE (2006) A UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination. *Molecular cell* 21(6):873-880.

56. Hirano S, *et al.* (2006) Structural basis of ubiquitin recognition by mammalian Eap45 GLUE domain. *Nat Struct Mol Biol* 13(11):1031-1032.

57. Schreiner P, *et al.* (2008) Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* 453(7194):548-552.

58. Swanson KA, Kang RS, Stamenova SD, Hicke L, & Radhakrishnan I (2003) Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. *The EMBO journal* 22(18):4597-4606.

59. Fisher RD, *et al.* (2003) Structure and ubiquitin binding of the ubiquitin-interacting motif. *The Journal of biological chemistry* 278(31):28976-28984.

60. Hofmann K & Falquet L (2001) A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends in biochemical sciences* 26(6):347-350.

61. Ohno A, *et al.* (2005) Structure of the UBA domain of Dsk2p in complex with ubiquitin molecular determinants for ubiquitin recognition. *Structure* 13(4):521-532.

62. Swanson KA, Hicke L, & Radhakrishnan I (2006) Structural basis for monoubiquitin recognition by the Ede1 UBA domain. *Journal of molecular biology* 358(3):713-724.

63. Burschowsky D, *et al.* (2011) Structural analysis of the conserved ubiquitin-binding motifs (UBMs) of the translesion polymerase iota in complex with ubiquitin. *The Journal of biological chemistry* 286(2):1364-1373.

64. Bomar MG, *et al.* (2010) Unconventional ubiquitin recognition by the ubiquitin-binding motif within the Y family DNA polymerases iota and Rev1. *Molecular cell* 37(3):408-417.

65. Prag G, *et al.* (2003) Mechanism of ubiquitin recognition by the CUE domain of Vps9p. *Cell* 113(5):609-620.

66. Mizuno E, Kawahata K, Kato M, Kitamura N, & Komada M (2003) STAM proteins bind ubiquitinated proteins on the early endosome via the VHS domain and ubiquitin-interacting motif. *Molecular biology of the cell* 14(9):3675-3689.

67. Rizzo AA, Salerno PE, Bezsonova I, & Korzhnev DM (2014) NMR structure of the human Rad18 zinc finger in complex with ubiquitin defines a class of UBZ domains in proteins linked to the DNA damage response. *Biochemistry* 53(37):5895-5906.

68. Bomar MG, Pai MT, Tzeng SR, Li SS, & Zhou P (2007) Structure of the ubiquitin-binding zinc finger domain of human DNA Y-polymerase eta. *EMBO reports* 8(3):247-251.

69. Cordier F, *et al.* (2009) The zinc finger of NEMO is a functional ubiquitin-binding domain. *The Journal of biological chemistry* 284(5):2902-2907.

70. Sundquist WI, *et al.* (2004) Ubiquitin recognition by the human TSG101 protein. *Molecular cell* 13(6):783-789.

71. VanDemark AP, Hofmann RM, Tsui C, Pickart CM, & Wolberger C (2001) Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. *Cell* 105(6):711-720.

72. Alam SL, *et al.* (2006) Structural basis for ubiquitin recognition by the human ESCRT-II EAP45 GLUE domain. *Nat Struct Mol Biol* 13(11):1029-1030.

73. Korzhnev DM, Bezsonova I, Lee S, Chalikian TV, & Kay LE (2009) Alternate binding modes for a ubiquitin-SH3 domain interaction studied by NMR spectroscopy. *Journal of molecular biology* 386(2):391-405.

74. Bezsonova I, *et al.* (2008) Interactions between the three CIN85 SH3 domains and ubiquitin: implications for CIN85 ubiquitination. *Biochemistry* 47(34):8937-8949.

75. He Y, Hicke L, & Radhakrishnan I (2007) Structural basis for ubiquitin recognition by SH3 domains. *Journal of molecular biology* 373(1):190-196.

76. Ortega Roldan JL, *et al.* (2013) Distinct ubiquitin binding modes exhibited by SH3 domains: molecular determinants and functional implications. *PloS one* 8(9):e73018.

77. Fu QS, *et al.* (2009) Structural basis for ubiquitin recognition by a novel domain from human phospholipase A2-activating protein. *The Journal of biological chemistry* 284(28):19043-19052.

78. Bellare P, Kutach AK, Rines AK, Guthrie C, & Sontheimer EJ (2006) Ubiquitin binding by a variant Jab1/MPN domain in the essential pre-mRNA splicing factor Prp8p. *Rna* 12(2):292-302.

79. Hofmann K (2009) Ubiquitin-binding domains and their role in the DNA damage response. *DNA repair* 8(4):544-556.

80. Andre I, Strauss CE, Kaplan DB, Bradley P, & Baker D (2008) Emergence of symmetry in homooligomeric biological assemblies. *Proceedings of the National Academy of Sciences of the United States of America* 105(42):16148-16152.

81. Goodsell DS & Olson AJ (2000) Structural symmetry and protein function. *Annual review of biophysics and biomolecular structure* 29:105-153.

82. Wolynes PG (1996) Symmetry and the energy landscapes of biomolecules. *Proceedings of the National Academy of Sciences of the United States of America* 93(25):14249-14255.

83. Sanders CR, 2nd, *et al.* (1996) Escherichia coli diacylglycerol kinase is an alpha-helical polytopic membrane protein and can spontaneously insert into preformed lipid vesicles. *Biochemistry* 35(26):8610-8618.

84. Hollenstein K, Dawson RJ, & Locher KP (2007) Structure and mechanism of ABC transporter proteins. *Current opinion in structural biology* 17(4):412-418.

85. Hollenstein K, Frei DC, & Locher KP (2007) Structure of an ABC transporter in complex with its binding protein. *Nature* 446(7132):213-216.

86. Bahar MW, Graham SC, Stuart DI, & Grimes JM (2011) Insights into the evolution of a complex virus from the crystal structure of vaccinia virus D13. *Structure* 19(7):1011-1020.

87. Clore GM, *et al.* (1995) Refined solution structure of the oligomerization domain of the tumour suppressor p53. *Nature structural biology* 2(4):321-333.

88. Van Horn WD, *et al.* (2009) Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase. *Science* 324(5935):1726-1729.

89. Li D & Caffrey M (2014) Renaturing membrane proteins in the lipid cubic phase, a nanoporous membrane mimetic. *Scientific reports* 4:5806.

90. Barette J, Velyvis A, Religa TL, Korzhnev DM, & Kay LE (2012) Cross-validation of the structure of a transiently formed and low populated FF domain folding intermediate determined by relaxation dispersion NMR and CS-Rosetta. *The journal of physical chemistry. B* 116(23):6637-6644.

91. Vogeli B, *et al.* (2014) Towards a true protein movie: a perspective on the potential impact of the ensemble-based structure determination using exact NOEs. *J Magn Reson* 241:53-59.

92. Kosen PA (1989) Spin labeling of proteins. *Methods in enzymology* 177:86-121.

93. Battiste JL & Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* 39(18):5355-5365.

94. Iwahara J, Anderson DE, Murphy EC, & Clore GM (2003) EDTA-derivatized deoxythymidine as a tool for rapid determination of protein binding polarity to DNA by intermolecular paramagnetic relaxation enhancement. *Journal of the American Chemical Society* 125(22):6634-6635.

95. Berliner LJ, Grunwald J, Hankovszky HO, & Hideg K (1982) A novel reversible thiol-specific spin label: papain active site labeling and inhibition. *Analytical biochemistry* 119(2):450-455.

96. Tjandra N & Bax A (1997) Measurement of dipolar contributions to 1JCH splittings from magnetic-field dependence of J modulation in two-dimensional NMR spectra. *J Magn Reson* 124(2):512-515.

97. Bax A, Kontaxis G, & Tjandra N (2001) Dipolar couplings in macromolecular structure determination. *Methods in enzymology* 339:127-174.

98. Prestegard JH, al-Hashimi HM, & Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys* 33(4):371-424.

99. Bax A & Tjandra N (1997) High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. *J Biomol NMR* 10(3):289-292.

100. Ottiger M & Bax A (1998) Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. *J Biomol NMR* 12(3):361-372.

101. Hansen MR, Mueller L, & Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature structural biology* 5(12):1065-1074.

102. Chou JJ, Gaemers S, Howder B, Louis JM, & Bax A (2001) A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *J Biomol NMR* 21(4):377-382.

103. Sass HJ, Musco G, Stahl SJ, Wingfield PT, & Grzesiek S (2000) Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J Biomol NMR* 18(4):303-309.

104. Cierpicki T & Bushweller JH (2004) Charged gels as orienting media for measurement of residual dipolar couplings in soluble and integral membrane proteins. *Journal of the American Chemical Society* 126(49):16259-16266.

105. Lorieau J, Yao L, & Bax A (2008) Liquid crystalline phase of G-tetrad DNA for NMR study of detergent-solubilized proteins. *Journal of the American Chemical Society* 130(24):7536-7537.

106. Meier S, Haussinger D, & Grzesiek S (2002) Charged acrylamide copolymer gels as media for weak alignment. *J Biomol NMR* 24(4):351-356.

107. Wang AC & Bax A (1995) Reparameterization of the Karplus Relation for $^{3}J(H^{\alpha}-N)$ and $^{3}J(H^{N}-C')$ in Peptides from Uniformly $^{13}C/^{15}N$-Enriched Human Ubiquitin. *Journal of the American Chemical Society* 117:1810-1813.

108. Shen Y, Delaglio F, Cornilescu G, & Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44(4):213-223.

109. Donald BR & Martin J (2009) Automated NMR Assignment and Protein Structure Determination using Sparse Dipolar Coupling Constraints. *Prog Nucl Magn Reson Spectrosc* 55(2):101-127.

110. Martin JW, Yan AK, Bailey-Kellogg C, Zhou P, & Donald BR (2011) A graphical method for analyzing distance restraints using residual dipolar couplings for structure determination of symmetric protein homo-oligomers. *Protein science : a publication of the Protein Society* 20(6):970-985.

111. Breeze AL (2000) Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Progress in Nuclear Magnetic Resonance Spectroscopy* 36:323-372.

112. Otting G & Wuthrich K (1990) Heteronuclear filters in two-dimensional [1H,1H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. *Q Rev Biophys* 23(1):39-96.

113. Zwahlen C, *et al.* (1997) Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage N-Peptide/boxB RNA Complex. *Journal of the American Chemical Society* 119:6711-6721.

114. Stuart AC, Borzilleri KA, Withka JM, & Palmer AG, 3rd (1999) Compensating for Variations in $^{1}H-^{13}C$ Scalar Coupling Constants in Isotope-Filtered NMR Experiments. *Journal of the American Chemical Society* 121:5346-5347.

115. Grzesiek S & Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR* 3(2):185-204.

116. Tjandra N & Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278(5340):1111-1114.

117. Valentine ER, Ferrage F, Massi F, Cowburn D, & Palmer AG, 3rd (2007) Joint composite-rotation adiabatic-sweep isotope filtration. *J Biomol NMR* 38(1):11-22.

118. Mikami S, *et al.* (2014) Structural insights into the recruitment of SMRT by the corepressor SHARP under phosphorylative regulation. *Structure* 22(1):35-46.

119. Brubaker K, *et al.* (2000) Solution structure of the interacting domains of the Mad-Sin3 complex: implications for recruitment of a chromatin-modifying complex. *Cell* 103(4):655-665.

120. Clore GM, *et al.* (1994) High-resolution structure of the oligomerization domain of p53 by multidimensional NMR. *Science* 265(5170):386-391.

121. Cho Y, Gorina S, Jeffrey PD, & Pavletich NP (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* 265(5170):346-355.

122. Clore GM, *et al.* (1995) Interhelical angles in the solution structure of the oligomerization domain of p53: correction. *Science* 267(5203):1515-1516.

123. Pervushin K (2000) Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q Rev Biophys* 33(2):161-197.

124. Felli IC & Brutscher B (2009) Recent advances in solution NMR: fast methods and heteronuclear direct detection. *Chemphyschem : a European journal of chemical physics and physical chemistry* 10(9-10):1356-1368.

125. Hyberts SG, Arthanari H, Robson SA, & Wagner G (2014) Perspectives in magnetic resonance: NMR in the post-FFT era. *J Magn Reson* 241:60-73.

126. Coggins BE & Zhou P (2007) Sampling of the NMR time domain along concentric rings. *J Magn Reson* 184(2):207-221.

127. Coggins BE, Venters RA, & Zhou P (2010) Radial sampling for fast NMR: Concepts and practices over three decades. *Prog Nucl Magn Reson Spectrosc* 57(4):381-419.

128. Coggins BE & Zhou P (2006) Polar Fourier transforms of radially sampled NMR data. *J Magn Reson* 182(1):84-95.

129. Kazimierczuk K, Kozminski W, & Zhukov I (2006) Two-dimensional Fourier transform of arbitrarily sampled NMR data sets. *J Magn Reson* 179(2):323-328.

130. Marion D (2006) Processing of ND NMR spectra sampled in polar coordinates: a simple Fourier transform instead of a reconstruction. *J Biomol NMR* 36(1):45-54.

131. Kupce E & Freeman R (2004) Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy. *Journal of the American Chemical Society* 126(20):6429-6440.

132. Kupce E & Freeman R (2003) Reconstruction of the three-dimensional NMR spectrum of a protein from a set of plane projections. *J Biomol NMR* 27(4):383-387.

133. Venters RA, Coggins BE, Kojetin D, Cavanagh J, & Zhou P (2005) (4,2)D Projection--reconstruction experiments for protein backbone assignment: application to human carbonic anhydrase II and calbindin D(28K). *Journal of the American Chemical Society* 127(24):8785-8795.

134. Coggins BE & Zhou P (2008) High resolution 4-D spectroscopy with sparse concentric shell sampling and FFT-CLEAN. *J Biomol NMR* 42(4):225-239.

135. Kazimierczuk K, Zawadzka A, Kozminski W, & Zhukov I (2006) Random sampling of evolution time space and Fourier transform processing. *J Biomol NMR* 36(3):157-168.

136. Ikura M, Kay LE, & Bax A (1990) A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* 29(19):4659-4667.

137. Grzesiek S, Kuboniwa H, Hinck AP, & Bax A (1995) Multiple-Quantum Line Narrowing for Measurement of H^{α}-H^{β} J Couplings in Isotopically Enriched Proteins. *Journal of the American Chemical Society* 117:5312-5315.

138. Bhat TN (1988) Calculation of an OMIT map. *J Appl Cryst* 21:279-281.

139. Nudelman I, Akabayov SR, Scherf T, & Anglister J (2011) Observation of intermolecular interactions in large protein complexes by 2D-double difference nuclear Overhauser enhancement spectroscopy: application to the 44 kDa interferon-receptor complex. *Journal of the American Chemical Society* 133(37):14755-14764.

140. Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, & Kozminski W (2010) Random sampling in multidimensional NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc* 57(4):420-434.

141. Coggins BE, Werner-Allen JW, Yan A, & Zhou P (2012) Rapid Protein Global Fold Determination Using Ultrasparse Sampling, High-Dynamic Range Artifact Suppression, and Time-Shared NOESY. *Journal of the American Chemical Society* 134:18619-18630.

142. Tao Y, Strelkov SV, Mesyanzhinov VV, & Rossmann MG (1997) Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure* 5(6):789-798.

143. Efimov VP, *et al.* (1994) Fibritin encoded by bacteriophage T4 gene wac has a parallel triple-stranded alpha-helical coiled-coil structure. *Journal of molecular biology* 242(4):470-486.

144. Letarov AV, Londer YY, Boudko SP, & Mesyanzhinov VV (1999) The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibritin. *Biochemistry. Biokhimiia* 64(7):817-823.

145. Guthe S, *et al.* (2004) Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. *Journal of molecular biology* 337(4):905-915.

146. Meier S, Guthe S, Kiefhaber T, & Grzesiek S (2004) Foldon, the natural trimerization domain of T4 fibritin, dissociates into a monomeric A-state form containing a stable beta-hairpin: atomic details of trimer dissociation and local beta-hairpin stability from residual dipolar couplings. *Journal of molecular biology* 344(4):1051-1069.

147. Reardon PN, *et al.* (2014) Structure of an HIV-1-neutralizing antibody target, the lipid-bound gp41 envelope membrane proximal region trimer. *Proceedings of the National Academy of Sciences of the United States of America* 111(4):1391-1396.

148. Lu Y, Welsh JP, & Swartz JR (2014) Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. *Proceedings of the National Academy of Sciences of the United States of America* 111(1):125-130.

149. Herrmann T, Guntert P, & Wuthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software

CANDID and the torsion angle dynamics algorithm DYANA. *Journal of molecular biology* 319(1):209-227.

150. Güntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278:353-378.

151. Delaglio F, *et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277-293.

152. Goddard TD & Kneller DG (2008) Sparky 3. (Univeristy of California, San Francisco).

153. Bartels C, Xia TH, Billeter M, Guntert P, & Wuthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* 6(1):1-10.

154. Waters LS, *et al.* (2009) Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiology and molecular biology reviews : MMBR* 73(1):134-154.

155. Saugar I, Ortiz-Bazan MA, & Tercero JA (2014) Tolerating DNA damage during eukaryotic chromosome replication. *Experimental cell research*.

156. Ohmori H, *et al.* (2001) The Y-family of DNA polymerases. *Molecular cell* 8(1):7-8.

157. Yang W (2014) An overview of Y-Family DNA polymerases and a case study of human DNA polymerase eta. *Biochemistry* 53(17):2793-2803.

158. Goodman MF & Woodgate R (2013) Translesion DNA polymerases. *Cold Spring Harbor perspectives in biology* 5(10):a010363.

159. Yang W & Woodgate R (2007) What a difference a decade makes: insights into translesion DNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 104(40):15591-15598.

160. Prakash S, Johnson RE, & Prakash L (2005) Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annual review of biochemistry* 74:317-353.

161. Ling H, Boudsocq F, Woodgate R, & Yang W (2001) Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* 107(1):91-102.

162. Schneider S, Schorr S, & Carell T (2009) Crystal structure analysis of DNA lesion repair and tolerance mechanisms. *Current opinion in structural biology* 19(1):87-95.

163. Yang W (2003) Damage repair DNA polymerases Y. *Current opinion in structural biology* 13(1):23-30.

164. Goodman MF (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annual review of biochemistry* 71:17-50.

165. Nair DT, Johnson RE, Prakash S, Prakash L, & Aggarwal AK (2004) Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing. *Nature* 430(6997):377-380.

166. Kunkel TA (2004) DNA replication fidelity. *The Journal of biological chemistry* 279(17):16895-16898.

167. Wojtaszek J, *et al.* (2012) Structural basis of Rev1-mediated assembly of a quaternary vertebrate translesion polymerase complex consisting of Rev1, heterodimeric polymerase (Pol) zeta, and Pol kappa. *The Journal of biological chemistry* 287(40):33836-33846.

168. Fischhaber PL & Friedberg EC (2005) How are specialized (low-fidelity) eukaryotic polymerases selected and switched with high-fidelity polymerases during translesion DNA synthesis? *DNA repair* 4(2):279-283.

169. Lehmann AR, *et al.* (2007) Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA repair* 6(7):891-899.

170. Friedberg EC, Lehmann AR, & Fuchs RP (2005) Trading places: how do DNA polymerases switch during translesion DNA synthesis? *Molecular cell* 18(5):499-505.

171. Hoege C, Pfander B, Moldovan GL, Pyrowolakis G, & Jentsch S (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* 419(6903):135-141.

172. Kannouche PL, Wing J, & Lehmann AR (2004) Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. *Molecular cell* 14(4):491-500.

173. Freudenthal BD, Gakhar L, Ramaswamy S, & Washington MT (2010) Structure of monoubiquitinated PCNA and implications for translesion synthesis and DNA polymerase exchange. *Nat Struct Mol Biol* 17(4):479-484.

174. Hishiki A, *et al.* (2009) Structural basis for novel interactions between human translesion synthesis polymerases and proliferating cell nuclear antigen. *The Journal of biological chemistry* 284(16):10552-10560.

175. Haracska L, *et al.* (2001) Targeting of human DNA polymerase iota to the replication machinery via interaction with PCNA. *Proceedings of the National Academy of Sciences of the United States of America* 98(25):14256-14261.

176. Haracska L, *et al.* (2001) Physical and functional interactions of human DNA polymerase eta with PCNA. *Molecular and cellular biology* 21(21):7199-7206.

177. Haracska L, *et al.* (2002) Stimulation of DNA synthesis activity of human DNA polymerase kappa by PCNA. *Molecular and cellular biology* 22(3):784-791.

178. Vidal AE, *et al.* (2004) Proliferating cell nuclear antigen-dependent coordination of the biological functions of human DNA polymerase iota. *The Journal of biological chemistry* 279(46):48360-48368.

179. Guo C, *et al.* (2006) REV1 protein interacts with PCNA: significance of the REV1 BRCT domain in vitro and in vivo. *Molecular cell* 23(2):265-271.

180. Livneh Z, Ziv O, & Shachar S (2010) Multiple two-polymerase mechanisms in mammalian translesion DNA synthesis. *Cell cycle* 9(4):729-735.

181. Yang K, Weinacht CP, & Zhuang Z (2013) Regulatory role of ubiquitin in eukaryotic DNA translesion synthesis. *Biochemistry* 52(19):3217-3228.

182. Kannouche PL & Lehmann AR (2004) Ubiquitination of PCNA and the polymerase switch in human cells. *Cell cycle* 3(8):1011-1013.

183. Brown S, Niimi A, & Lehmann AR (2009) Ubiquitination and deubiquitination of PCNA in response to stalling of the replication fork. *Cell cycle* 8(5):689-692.

184. Davies AA, Huttner D, Daigaku Y, Chen S, & Ulrich HD (2008) Activation of ubiquitin-dependent DNA damage bypass is mediated by replication protein a. *Molecular cell* 29(5):625-636.

185. Ulrich HD & Jentsch S (2000) Two RING finger proteins mediate cooperation between ubiquitin-conjugating enzymes in DNA repair. *The EMBO journal* 19(13):3388-3397.

186. Brusky J, Zhu Y, & Xiao W (2000) UBC13, a DNA-damage-inducible gene, is a member of the error-free postreplication repair pathway in Saccharomyces cerevisiae. *Current genetics* 37(3):168-174.

187. Hofmann RM & Pickart CM (1999) Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. *Cell* 96(5):645-653.

188. Broomfield S, Chow BL, & Xiao W (1998) MMS2, encoding a ubiquitin-conjugating-enzyme-like protein, is a member of the yeast error-free postreplication repair pathway. *Proceedings of the National Academy of Sciences of the United States of America* 95(10):5678-5683.

189. Chiu RK, *et al.* (2006) Lysine 63-polyubiquitination guards against translesion synthesis-induced mutations. *PLoS genetics* 2(7):e116.

190. Chang DJ & Cimprich KA (2009) DNA damage tolerance: when it's OK to make mistakes. *Nature chemical biology* 5(2):82-90.

191. Xiao W, Chow BL, Broomfield S, & Hanna M (2000) The Saccharomyces cerevisiae RAD6 group is composed of an error-prone and two error-free postreplication repair pathways. *Genetics* 155(4):1633-1641.

192. Motegi A, *et al.* (2006) Human SHPRH suppresses genomic instability through proliferating cell nuclear antigen polyubiquitination. *The Journal of cell biology* 175(5):703-708.

193. Brun J, *et al.* (2008) hMMS2 serves a redundant role in human PCNA polyubiquitination. *BMC molecular biology* 9:24.

194. Huang TT, *et al.* (2006) Regulation of monoubiquitinated PCNA by DUB autocleavage. *Nature cell biology* 8(4):339-347.

195. Guo C, *et al.* (2006) Ubiquitin-binding motifs in REV1 protein are required for its role in the tolerance of DNA damage. *Molecular and cellular biology* 26(23):8892-8900.

196. Parker JL, Bielen AB, Dikic I, & Ulrich HD (2007) Contributions of ubiquitin- and PCNA-binding domains to the activity of Polymerase eta in Saccharomyces cerevisiae. *Nucleic acids research* 35(3):881-889.

197. Bienko M, *et al.* (2010) Regulation of translesion synthesis DNA polymerase eta by monoubiquitination. *Molecular cell* 37(3):396-407.

198. Kim H, Yang KL, Dejsuphong D, & D'Andrea AD (2012) Regulation of Rev1 by the Fanconi anemia core complex. *Nat Struct Mol Biol* 19(2):164-170.

199. Watanabe K, *et al.* (2004) Rad18 guides poleta to replication stalling sites through physical interaction and PCNA monoubiquitination. *The EMBO journal* 23(19):3886-3896.

200. Plosky BS, *et al.* (2006) Controlling the subcellular localization of DNA polymerases iota and eta via interactions with ubiquitin. *The EMBO journal* 25(12):2847-2855.

201. Guo C, Tang TS, Bienko M, Dikic I, & Friedberg EC (2008) Requirements for the interaction of mouse Polkappa with ubiquitin and its biological significance. *The Journal of biological chemistry* 283(8):4658-4664.

202. McIntyre J, *et al.* (2013) Ubiquitin mediates the physical and functional interaction between human DNA polymerases eta and iota. *Nucleic acids research* 41(3):1649-1660.

203. Sabbioneda S, *et al.* (2008) Effect of proliferating cell nuclear antigen ubiquitination and chromatin structure on the dynamic properties of the Y-family DNA polymerases. *Molecular biology of the cell* 19(12):5193-5202.

204. Zhuang Z, *et al.* (2008) Regulation of polymerase exchange between Poleta and Poldelta by monoubiquitination of PCNA and the movement of DNA polymerase holoenzyme. *Proceedings of the National Academy of Sciences of the United States of America* 105(14):5361-5366.

205. Wood A, Garg P, & Burgers PM (2007) A ubiquitin-binding motif in the translesion DNA polymerase Rev1 mediates its essential functional interaction with ubiquitinated proliferating cell nuclear antigen in response to DNA damage. *The Journal of biological chemistry* 282(28):20256-20263.

206. D'Souza S, Waters LS, & Walker GC (2008) Novel conserved motifs in Rev1 C-terminus are required for mutagenic DNA damage tolerance. *DNA repair* 7(9):1455-1470.

207. Woodruff RV, Bomar MG, D'Souza S, Zhou P, & Walker GC (2010) The unusual UBZ domain of Saccharomyces cerevisiae polymerase eta. *DNA repair* 9(11):1130-1141.

208. Ai Y, *et al.* (2011) A novel ubiquitin binding mode in the S. cerevisiae translesion synthesis DNA polymerase eta. *Molecular bioSystems* 7(6):1874-1882.

209. Tissier A, McDonald JP, Frank EG, & Woodgate R (2000) poliota, a remarkably error-prone human DNA polymerase. *Genes Dev* 14(13):1642-1650.

210. Johnson RE, Washington MT, Haracska L, Prakash S, & Prakash L (2000) Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. *Nature* 406(6799):1015-1019.

211. Zhang Y, Yuan F, Wu X, & Wang Z (2000) Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase iota. *Molecular and cellular biology* 20(19):7099-7108.

212. Petta TB, *et al.* (2008) Human DNA polymerase iota protects cells against oxidative stress. *The EMBO journal* 27(21):2883-2895.

213. Kannouche P, *et al.* (2003) Localization of DNA polymerases eta and iota to the replication machinery is tightly co-ordinated in human cells. *The EMBO journal* 22(5):1223-1233.

214. McDonald JP, *et al.* (2003) 129-derived strains of mice are deficient in DNA polymerase iota and have normal immunoglobulin hypermutation. *The Journal of experimental medicine* 198(4):635-643.

215. Kay LE, Torchia DA, & Bax A (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28(23):8972-8979.

216. LeMaster DM & Richards FM (1988) NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteriation. *Biochemistry* 27(1):142-150.

217. Weisemann R, Ruterjans H, & Bermel W (1993) 3D triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins. *J Biomol NMR* 3(1):113-120.

218. Richardson JS & Richardson DC (1988) Amino acid preferences for specific locations at the ends of alpha helices. *Science* 240(4859):1648-1652.

219. Davis IW, *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research* 35(Web Server issue):W375-383.

220. Chen ZJ & Sun LJ (2009) Nonproteolytic functions of ubiquitin in cell signaling. *Molecular cell* 33(3):275-286.

221. Chang YG, *et al.* (2006) Solution structure of the ubiquitin-associated domain of human BMSC-UbP and its complex with ubiquitin. *Protein science : a publication of the Protein Society* 15(6):1248-1259.

222. Prag G, *et al.* (2005) Structural mechanism for ubiquitinated-cargo recognition by the Golgi-localized, gamma-ear-containing, ADP-ribosylation-factor-binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* 102(7):2334-2339.

223. Guo C, *et al.* (2003) Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. *The EMBO journal* 22(24):6621-6630.

224. Crosetto N, *et al.* (2008) Human Wrnip1 is localized in replication factories in a ubiquitin-binding zinc finger-dependent manner. *The Journal of biological chemistry* 283(50):35173-35185.

225. Zhou P, Lugovskoy AA, & Wagner G (2001) A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. *J Biomol NMR* 20(1):11-14.

226. Shapiro RA, Brindley AJ, & Martin RW (2010) Thermal stabilization of DMPC/DHPC bicelles by addition of cholesterol sulfate. *Journal of the American Chemical Society* 132(33):11406-11407.

227. Ball G, *et al.* (2006) Measurement of one-bond 13Calpha-1Halpha residual dipolar coupling constants in proteins by selective manipulation of CalphaHalpha spins. *J Magn Reson* 180(1):127-136.

228. Rosenberg PS, Tamary H, & Alter BP (2011) How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi Anemia in the United States and Israel. *American journal of medical genetics. Part A* 155A(8):1877-1883.

229. D'Andrea AD & Grompe M (2003) The Fanconi anaemia/BRCA pathway. *Nature reviews. Cancer* 3(1):23-34.

230. de Winter JP & Joenje H (2009) The genetic and molecular basis of Fanconi anemia. *Mutation research* 668(1-2):11-19.

231. Kennedy RD & D'Andrea AD (2005) The Fanconi Anemia/BRCA pathway: new faces in the crowd. *Genes Dev* 19(24):2925-2940.

232. Rego MA, Kolling FWt, & Howlett NG (2009) The Fanconi anemia protein interaction network: casting a wide net. *Mutation research* 668(1-2):27-41.

233. Thompson LH, Hinz JM, Yamada NA, & Jones NJ (2005) How Fanconi anemia proteins promote the four Rs: replication, recombination, repair, and recovery. *Environ Mol Mutagen* 45(2-3):128-142.

234. Wang W (2007) Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nature reviews. Genetics* 8(10):735-748.

235. Kozekov ID, *et al.* (2003) DNA interchain cross-links formed by acrolein and crotonaldehyde. *Journal of the American Chemical Society* 125(1):50-61.

236. Auerbach AD (1988) A test for Fanconi's anemia. *Blood* 72(1):366-367.

237. German J, *et al.* (1987) A test for Fanconi's anemia. *Blood* 69(6):1637-1641.

238. Sasaki MS (1975) Is Fanconi's anaemia defective in a process essential to the repair of DNA cross links? *Nature* 257(5526):501-503.

239. Scharer OD (2005) DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem : a European journal of chemical biology* 6(1):27-32.

240. Walden H & Deans AJ (2014) The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annual review of biophysics* 43:257-278.

241. Moldovan GL & D'Andrea AD (2009) How the fanconi anemia pathway guards the genome. *Annual review of genetics* 43:223-249.

242. Kim H & D'Andrea AD (2012) Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes Dev* 26(13):1393-1408.

243. Collis SJ, *et al.* (2008) FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex. *Molecular cell* 32(3):313-324.

244. Ciccia A, *et al.* (2007) Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Molecular cell* 25(3):331-343.

245. Singh TR, *et al.* (2010) MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM. *Molecular cell* 37(6):879-886.

246. Pulliam-Leath AC, *et al.* (2010) Genetic disruption of both Fancc and Fancg in mice recapitulates the hematopoietic manifestations of Fanconi anemia. *Blood* 116(16):2915-2920.

247. Cole AR, Lewis LP, & Walden H (2010) The structure of the catalytic subunit FANCL of the Fanconi anemia core complex. *Nat Struct Mol Biol* 17(3):294-298.

248. Garner E & Smogorzewska A (2011) Ubiquitylation and the Fanconi anemia pathway. *FEBS letters* 585(18):2853-2860.

249. Crossan GP & Patel KJ (2012) The Fanconi anaemia pathway orchestrates incisions at sites of crosslinked DNA. *The Journal of pathology* 226(2):326-337.

250. Huang M & D'Andrea AD (2010) A new nuclease member of the FAN club. *Nat Struct Mol Biol* 17(8):926-928.

251. Yamamoto KN, *et al.* (2011) Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. *Proceedings of the National Academy of Sciences of the United States of America* 108(16):6492-6496.

252. Ciccia A, McDonald N, & West SC (2008) Structural and functional relationships of the XPF/MUS81 family of proteins. *Annual review of biochemistry* 77:259-287.

253. Nijman SM, *et al.* (2005) The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Molecular cell* 17(3):331-339.

254. Cohn MA, *et al.* (2007) A UAF1-containing multisubunit protein complex regulates the Fanconi anemia pathway. *Molecular cell* 28(5):786-797.

255. Ali AM, *et al.* (2012) FAAP20: a novel ubiquitin-binding FA nuclear core-complex protein required for functional integrity of the FA-BRCA DNA repair pathway. *Blood* 119(14):3285-3294.

256. Leung JW, *et al.* (2012) Fanconi anemia (FA) binding protein FAAP20 stabilizes FA complementation group A (FANCA) and participates in interstrand cross-link repair. *Proceedings of the National Academy of Sciences of the United States of America* 109(12):4491-4496.

257. Kim H, Yang K, Dejsuphong D, & D'Andrea AD (2012) Regulation of Rev1 by the Fanconi anemia core complex. *Nat Struct Mol Biol* 19(2):164-170.

258. Yan Z, *et al.* (2012) A ubiquitin-binding protein, FAAP20, links RNF8-mediated ubiquitination to the Fanconi anemia DNA repair network. *Molecular cell* 47(1):61-75.

259. Iha H, *et al.* (2008) Inflammatory cardiac valvulitis in TAX1BP1-deficient mice through selective NF-kappaB activation. *The EMBO journal* 27(4):629-641.

260. Ogi T, Kannouche P, & Lehmann AR (2005) Localisation of human Y-family DNA polymerase kappa: relationship to PCNA foci. *Journal of cell science* 118(Pt 1):129-136.

261. Bish RA & Myers MP (2007) Werner helicase-interacting protein 1 binds polyubiquitin via its zinc finger domain. *The Journal of biological chemistry* 282(32):23184-23193.

262. Yang K, Moldovan GL, & D'Andrea AD (2010) RAD18-dependent recruitment of SNM1A to DNA repair complexes by a ubiquitin-binding zinc finger. *The Journal of biological chemistry* 285(25):19085-19091.

263. Cybulski KE & Howlett NG (2011) FANCP/SLX4: a Swiss army knife of DNA interstrand crosslink repair. *Cell cycle* 10(11):1757-1763.

264. Rizzo AA, Salerno PE, Bezsonova I, & Korzhnev DM (2014) NMR structure of the human Rad18 zinc finger in complex with ubiquitin defines a class of UBZ domains in proteins linked to the DNA damage response. *Biochemistry*.

265. Ceregido MA, *et al.* (2014) The structure of TAX1BP1 UBZ1+2 provides insight into target specificity and adaptability. *Journal of molecular biology* 426(3):674-690.

266. Wang S & Zhou P (2014) Sparsely-sampled, high-resolution 4-D omit spectra for detection and assignment of intermolecular NOEs of protein complexes. *J Biomol NMR* 59(2):51-56.

267. Werner-Allen JW, Jiang L, & Zhou P (2006) A 'just-in-time' HN(CA)CO experiment for the backbone assignment of large proteins with high sensitivity. *J Magn Reson* 181(1):177-180.

268. Wen J, Zhou P, & Wu J (2012) Efficient acquisition of high-resolution 4-D diagonal-suppressed methyl-methyl NOESY for large proteins. *J Magn Reson* 218:128-132.

Biography

Su Wang was born on Sep 28, 1985 in Yibin, Sichuan, China. After graduating from Yibin No.1 High School in 2004, she went to University of Science and Technology of China in Hefei, Anhui, China. In 2008, she earned her Bachelor of Science degree in Biological Science. In the fall of 2008, Su entered the Duke University Biochemistry Department's graduate program. She joined Pei Zhou's lab in the spring of 2009.

Publications:

1. Wojtaszek JL*, **Wang S***, Kim H, Wu Q, D'Andrea AD, Zhou P. Ubiquitin recognition by FAAP20 expands the complex interface beyond the canonical UBZ domain. Nuclear Acid Research 2014 (in press). (* equal contributions)

2. **Wang S** and Zhou P. Sparsely-sampled, high-resolution 4-D omit spectra for detection and assignment of intermolecular NOEs of protein complexes. J Biomol NMR 2014; 59: 51-6

3. Wojtaszek J, Liu J, D'Souza S, **Wang S**, Xue Y, Walker GC, Zhou P. Multifaceted recognition of vertebrate Rev1 by translesion polymerases zeta and kappa. J Biol Chem. 2012 July 27; 287 (31): 26400-8.

4. Werner-Allen JW, Lee CJ, Liu P, Nicely NI, **Wang S**, Greenleaf AL, Zhou P. Cis proline-mediated pSer5-dephosphorylation by the RNA polymerase II CTD phosphatase Ssu72. J Biol Chem. 286, 5717-5826 (2011).